GHRSST-PP

GODAE High Resolution Sea Surface Temperature Pilot Project

The Recommended GHRSST-PP Processing Specification (Version 1)

Reference: GHRSST/17
Issue: V0.8.32 DRAFT. To be finalised at the GHRSST-PP Science Team workshop, September 2003
Date of issue: Tuesday, 10 June 2003
Document type: Reference document
Author: The GHRSST-PP Science Team and ISDI-TAG working group

Document change record

Title:

The Recommended GHRSST-PP In situ and Satellite Integration Processing Model Version 1 (GDS-v1)

GHRSST-PP reference document number:

GHRSST/17

Approved by:

	Entity	Signature.	Date
G Wick	For the ISDI-TAG		
C J Donlon	For the GHRSST-PP Science Team		

Document change record

Author	Modification	Issue	Rev.	Date
C Donlon	Basic v3.0 complete.	DRAFT	V0.1	10th September 2002
C Donlon	Medspiration 1.5d URD and GDIP 1.2. now merged into GDS following discussions with ESA at GHRSST-III. Included significant changes based on the GHRSST-PP workshop which resulted in a complete re write and specification of the GDS.	DRAFT	V0.2	22nd December 2002
C Donlon	Added J. Vasquez and E. Armstrong comments on data management, MMR and data exchange. Comments from K Casey on data formats and QC	DRAFT	0.3	31st January 2003
C Donlon	Added general comments from I. Robinson	DRAFT	0.3.1	8th February 2003
C Donlon	General inputs from data providers added following January blanket mail request for comments	DRAFT	0.3.5	10th February 2003
C Donlon	Added section on diagnostics from R. Reynolds	DRAFT	0.4	15th February 2003
C Donlon	Included comments and discussions following GHRSST-PP seminar at the Met Office, UK.	DRAFT	0.4.1	19th February, 2003
C Donlon	Added contributions from P. Minnett and B. Evans for validation and WP-ID7	DRAFT	0.5	21st February 2003
C Donlon	Included comments of F. Blanc, P. LeTraon based on MERSEA-IP discussions for data management and the GODAE data sharing project	DRAFT	0.5.1	23rd February 2003
C Donlon	Added discussions from the International GODAE Steering Committee meeting, EUMETSAT, 17-21st Feb 2003	DRAF	0.5.2	26th February 2003
C Donlon	Added comments from M. Lynch (Curtin University, Perth, Australia) following 1 day discussion of GHRSST-PP at JRC on 28th Feb 2003	DRAFT	0.5.3	28th February, 2003
C Donlon	Added general review from P. LeBorgne comments	DRAFT	0.5.7	1 March 2003
C Donlon	Added additional P. LeBorgne comments following extensive discussions.	DRAFT	0.6	11th March 2003
C. Donlon	Included ideas and comments from various people at Met Office UK. (M. Bell (FOAM), R. Saunders (NWP satellite group)).	DRAFT	0.6.1	17th March 2003
C Donlon	Included comments from N. Rayner following meeting at Hedley Centre for Climate Change.	DRAFT	0.6.2	26th March 2003
C Donlon	Included comments following telephone discussion with G. Wick on merging procedures, grid size and general GDSv1 structure.	DRAFT	0.61	28th March -2003
C Donlon	Discussions with C. Mutlow on data products, reference processor and services included.	DRAFT	0.7.1	29th March 2003
C Donlon	Added GHRSST-PP output grid specifications after extensive discussions of all Science Team	DRAFT	0.72	9th April 2003
C Donlon	Included collective comments and discussions following EGS/AGU/EGU Oral presentation, Nice.	DRAFT	0.7.3	12th April 2003
C Donlon	Included NGSSTv1.5 material from H. Kawamura	DRAFT	0.7.4	13th April 2003
C Donlon	Included discussion (at EGS Nice) with S. Pouliquen (IFREMER) for CORIOLIS and MEDS joint service support.	DRAFT	0.7.5	14th April 2003

C Donlon	General review and edit	DRAFT	0.7.6	15th April 2003
C Donlon	Included new specifications for Diurnal Confidence flag generation based on inputs and discussion with A. Stuart-Menteth	DRAFT	0.8.0	9th May 2003
C Donlon	Included comments from GHRSST-PP Science team. Substantial revisions to simplify the processing of data., Renamed to a processing specification. Dropped L3C and L3M data products form the specification. Revised the Functional breakdown diagrams and enhanced the section on MDB and validation. L2P data products now contain diurnal variation information	DRAFT	0.8.1	21st May 2003
C Donlon	Included revisions based on input from D. May	DRAFT	0.8.2	25th May 2003
C Donlon	Included detailed comments from N Rayner	DRAFT	0.8.22	26th May 2003
C Donlon	Included detailed comments from Jim Cummings and K Casey	DRAFT	0.8.3	27th May 2003
C Donlon	Included detailed comments from Bill Emery, R. Reynolds, I Robinson, P. LeBorgne, A. Stuart-Menteth, P. Minnett.	DRAFT	0.8.31	28th May 2003
C Donlon	Additional comments by A. Stuart-Menteth	DRAFT	0.8.32	30th May 2003
C Donlon	Comments by A Stuart Menteth	DRAFT	0.8.32	3rd June 2003

Table of Contents

Executive Summary

The Global Ocean Data Assimilation Experiment (GODAE) High Resolution Sea Surface Temperature Pilot Project (GHRSST-PP) Development and Implementation Plan (GDIP) is based on a regional task sharing approach to the production and dissemination of GHRSST-PP data products. The GDIP describes a distributed data processing system that shares the data processing tasks that are necessary to operationally generate and distribute high resolution SST data sets having global coverage. In this model, Regional Data Assembly Centres (RDAC) generate regional coverage sea surface temperature (SST) data streams in real-time based on readily available processed SST geophysical data products (Level-2 data) and other satellite and in situ data streams. The L2 data are provided by data providers that in most cases will have already completed quality control procedures during the production of L2 SST products. RDAC data streams are then assembled together at Global Data Analysis Centres (GDAC) where they are integrated and analysed to provide a new generation of global coverage data products. GHRSST-PP users may access data products at either GDAC or RDAC in real time.

The GHRSST-PP Processing specification (GDS) is a common data processing specification that is implemented at each GHRSST-PP RDAC and GDAC. It represents a consensus opinion of the GHRSST-PP community of how to combine satellite and in situ data streams within a globally distributed operational system and provide a new generation of global coverage SST data products.

The GDS describes in detail the input and output data specifications, data processing procedures, algorithms and data product file formats **that are common to each GHRSST-PP RDAC and GDAC** in order for the GHRSST-PP shared implementation model to function efficiently. It recognises that RDAC and GDAC must implement specific data processing procedures that account for data streams that are regionally specific (e.g., geostationary imagers). Furthermore, it is expected that RDAC may provide additional data products and services that satisfy regional user requirements (e.g., regionally specific analysed data products or ultra-high resolution data products). Such regional extensions to the GDS are not included in the GDS. Instead, the GDS documents the minimum data processing that should be completed at each RDAC and GDAC in order that data products generated independently at different GDAC and RDAC can be used both individually and together with confidence. In order to verify that each processing centre has implemented the GDS correctly a reference GDS data processor will e used to process a set of data products using a GDS test data set. Each RDAC and GDAC should be able to provide the same output as the reference processor using the test data set as an input.

This document describes in detail the GDS version 1 (GDSv1) which is based on the First Report of the GHRSST-PP ISDI-TAG (Wick et al, 2002) and subsequent discussions at the Third GHRSST-PP Workshop, held at ESA/ESRIN, Frascati, Italy in December 2002 (Donlon et al., 2003b). It is a working reference document written primarily for the RDAC and GDAC teams as they physically realise, develop and, refine the GHRSST-PP demonstration system. As such, it may be considered a technical reference manual for the GHRSST-PP.

This first version of the GDS (GDSv1) is focused on an initial data processing model with specific emphasis on implementing

(a) An operational data exchange and delivery system between data providers, RDAC, GDAC and user communities by 2005,
(b) A first version of analysed global coverage SST data products based on current knowledge that can be upgraded and refined
(c) A suite of initial test data products to be available to selected users at each RDAC and GDAC by mid 2004.

As the GDSv1 is a prototype operational system, a program of on-going system evaluation, validation and assessment will be necessary to commission and validate the GDS in real time.

Much of the GDSv1 is dedicated to issues of data exchange, management and operational considerations. The GDS will evolve throughout this process and a significant scientific upgrade of the processor is foreseen following the successful commission of the GDSv1. Targeted upgrade paths are therefore clearly identified in a dedicated section of this document.

1 Introduction

The primary aim of the Global Ocean Data Assimilation Experiment (GODAE) High Resolution Sea Surface Temperature Pilot Project (GHRSST-PP) is to develop and operate an operational demonstration system that will deliver high-resolution (better than 10 km and ~ 6 hourly) global coverage SST data products for the diverse needs of GODAE and the wider scientific community. A new generation of global coverage SST data products will be derived and served to the user community by combining complementary Level-2 (L2) satellite and in situ observations in real time (6 hourly). A full description of the GHRSST-PP project is given in the GHRSST-PP Development and Implementation Plan (GDIP) which can be obtained from the GHRSST-PP project web server located at http://www.ghrsst-pp.org.

Figure 1.1 provides a simplified schematic overview of the GDIP. The GHRSST-PP is based on a distributed system in which the data processing operations that are necessary to operationally generate and distribute high resolution SST data sets having global coverage are shared by Regional Data Assembly Centres (RDAC). RDAC ingest, quality control and merge existing L2 satellite and in situ data sources that are then used together to generate regional coverage L3 SST data streams to the same specification, in real-time. RDAC data streams are then assembled together at Global Data Analysis Centres (GDAC) where they are integrated and analysed to provide L4 global coverage data products.

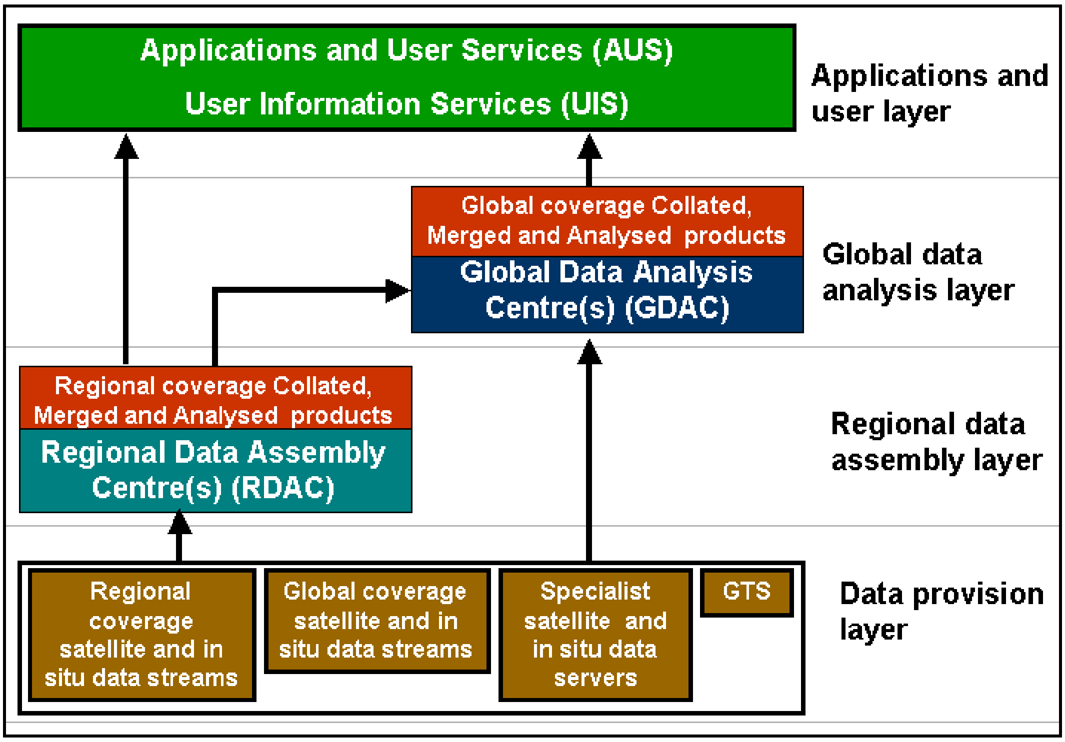


Figure 1.1. A Simplified schematic overview of the GHRSST-PP Development and Implementation Plan (GDIP) framework, identifying the major GHRSST-PP components and services. Red boxes indicate the output of GHRSST-PP data products.

Each RDAC has an associated project that is regionally funded to implement a processing model that conforms to the minimum specification described in the GDS. Five GHRSST-PP RDAC projects are currently in preparation and are fully described in the GDIP. These are:

1. The New Generations SST project (NGSST) serving the western Pacific area approximately defined by the Geostationary Meteorological Satellite (GMS), footprint. The NGSST project is based in Japan.
2. The European Medspiration Project (MSP) serving the Atlantic area.
3. The Ocean Modelling and Predication system (OMPS) serving the regional needs of the Australian and Indian Ocean region. The OMPS project is based in Australia.
4. The Survey of the environment Assisted by Satellite (SEASnet) program of the IDD serving the tropical oceans. The SEASnet project is based in France.
5. The Production of Enhanced Multi-sensor SST analysis project (PEMSA, funding TBC) serving the Eastern Pacific and Western Atlantic areas. The PEMSA project is based in the USA.

A single GDAC centre is in preparation at the Jet propulsion Laboratory Physical Oceanography Data Active Archive Center (PO.DAAC) which is linked to the US-GODAE data server system at Monterey (http://www.usgodae.org). The US-GODAE server is dedicated to serving large data sets to operational ocean models. In addition, the extensive experience and capability of the PO.DAAC will ensure that GHRSSTPP scientific users are provided with excellent support, documentation and data access.

The GHRSST-PP Processing Specification (GDS)

The GHRSST-PP Processing Specification is a recommended **common** data processing specification that should be implemented at each GHRSST-PP RDAC and GDAC. It represents a consensus opinion of the GHRSST-PP community of how satellite and in situ data streams should be combined within a globally distributed operational system to provide a new generation of global coverage SST data products.

The GDS defines clearly the input and output data specifications, data processing procedures, algorithms and data product file formats **that are common to each GHRSST-PP RDAC and GDAC** in order for the GHRSST-PP shared implementation model to function efficiently. It specifies in detail the minimum data processing that should be performed at each RDAC and GDAC centre to ensure that data and data products can be used and exchanged with confidence. For example, a common processing model is necessary to simplify documentation of data, facilitate exchange by sharing a common data format agreed by RDAC, GDAC and users, to avoid significant duplication of effort, to minimise reformatting of different data products derived by RDAC and to ease the integration of RDAC data to provide global coverage data sets at GDAC centres. The GDS will also provide low-level data products to the GHRSST-PP reanalysis project (RAN) that will improve on real time GDS data products by using additional data that are only available in a delayed mode and extensive quality control procedures. The GHRSST-PP RAN Strategy is described in Casey et al. (2003).

GHRSST-PP RDAC centres must implement data processing procedures that account for specific aspects of regionally specific data streams (e.g., geostationary imagers). RDAC must also provide additional data products and services to satisfy regional user requirements (e.g., regionally specific analysed data products or ultra-high resolution data products) that will require research and development for new analysis, quality control, and data provision procedures. These regional extensions to the GDS are critical to the preservation of regional identity and the proper evolution of the GDS; without evolution, the data products and services provided by the GHRSST-PP are unlikely to satisfy user demands and concerns. Clearly it is not possible to specify exactly how the final version of the GDS will be implemented and configured without first having a system in place to make an informed assessment. Neither is it desirable to have a GDS that is rigid without the ability to innovate based on experience. Therefore, this first version of the GDS (GDSv1) is focused on an initial data processing model with specific emphasis on implementing:

(a) An operational data exchange and delivery system between data providers, RDAC, GDAC and user communities by 2005,
(b) A first version of L4 data products based on current knowledge that can be upgraded and refined based on experience
(c) A suite of initial test data products to be available to selected users at each RDAC and GDAC by the end of 2004.

Document Scope

This document describes in detail the version 1 GDS (GDSv1) which is based on the First Report of the GHRSST-PP ISDI-TAG (Wick et al, 2002) and many subsequent discussions at the Third GHRSST-PP Workshop, held at ESA/ESRIN, Frascati, Italy in December 2002 (Donlon, 2003b). It represents a consensus opinion of the GHRSST-PP community of how to combine satellite and in situ data streams within a globally distributed operational system to provide a new generation of global coverage SST data products. Much of the document is dedicated to issues of data exchange, management and operational considerations. The GDSv1 will evolve and a significant scientific upgrade of the processor is foreseen following the successful commission of the GDSv1. Targeted upgrade paths are clearly identified in WP-ID9 of the GDSv1. It is a working document written primarily for the RDAC and GDAC teams as they physically realise, develop and, refine the GHRSST-PP demonstration system. As such, it may be considered a technical reference manual for the GHRSST-PP.

The GDS is split into several work package (WP) sections that are supported by extensive technical Appendices. This format provides a framework that preserves the readability of the processing specification while providing extensive technical references that can be easily maintained and updates without affecting the overall structure of the GDS. WP have been used extensively to develop a modular approach with clearly defined input and output parameters that greatly assist in the development of large multi-institute projects such as GHRSST-PP. While the interface parameters for each WP are fixed, considerable flexibility within a WP is maintained by this type of approach.

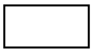
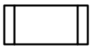
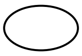
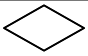
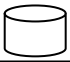
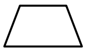
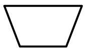
2 Notations, conventions and definitions used by the GDSv1.

The following notations and conventions are used throughout the GDS.

2.1 Flow diagram symbols

The symbols described in Table 2.1.1 are used in all flow/functional breakdown diagrams.

Table 2.1.1 Symbols used in GDS flow and functional breakdown diagrams.

Symbol	Meaning
	denotes an algorithm step
	denotes an algorithm step for which a further breakdown exists
	denotes a parameter
	denotes an interface parameter
	denotes a decision step
	denotes a data base
	denotes the start of a loop
	denotes the end of a loop

2.2 Definition of data processing levels

The GDSv1 uses the definitions provided in Table 2.2.2 when referring to data processing levels.

Table 2.2.1 Definition of satellite data processing levels.

Level	Abbreviation	Description
Level 0	L0	Unprocessed instrument and payload data at full resolution.
Level 1A	L1A	Reconstructed unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and geo-referencing parameters, computed and appended, but not applied, to the Level 0 data.
Level 1B	L1B	Level 1A data that have been processed to sensor units.
Level 2	L2	Geophysical variables derived from Level 1 source data at the same resolution and location as the Level 1 data.
Level 3	L3	Level 2 variables mapped on uniform space-time grid scales.
Level 4	L4	Results from analyses of lower level data (e.g., variables derived from multiple measurements).

2.3 GDSv1 data processing window specifications.

GDS data processing activities are linked to four processing window (PW) time periods within each 24 hour period. Each PW is defined in Table 2.3.1.

Table 2.3.1 Definition of GDSv1 data processing windows (PW).

Name	Temporal coverage/Eligible input data	Nominal output time	Latest actual output time
PW1	00:00 - 05: 59UTC	03:00 UTC	09:00 UTC
PW2	06:00 - 11:59 UTC	09:00 UTC	15:00 UTC
PW3	12:00 - 17:59 UTC	15:00 UTC	21:00 UTC
PW4	18:00 - 23:59 UTC	21:00 UTC	03:00 UTC (T+1)

Only data collected within the period are eligible for inclusion within a given PW. GDSv1 data products are only valid for a given PW within 3 hours of the PW eligible input data time (the latest actual output time) after which the following PW data products may be more representative.

GDS-v1 GDAC data processing activities are linked to Analysed Product Processing Window (APPW) periods within a 24 hour period that are defined in Table 2.3.2.

Table 2.3.2 Definition of GDSv1 GDAC analysed product processing windows (APPW).

Name	Temporal coverage/Eligible input data	Nominal output time	Latest actual output time
APPW1	00:00- 23:59 UTC	00:00 UTC	06:00 (T+1)

Only data collected within the APPW period are eligible for inclusion within a given APPW. Data products are only valid for a given APPW within 6 hours of the APPW output time after which the following APPW data products may be more representative.

2.4 GHRSST-PP definitions of sea surface temperature

Definitions of SST provide a necessary theoretical framework that can be used to understand the information content and relationships between measurements of SST made by different satellite and in situ instruments. Figure 2.4.1 presents a schematic diagram that provides this framework. It encapsulates the effects of dominant heat transport processes and time scales of variability associated with distinct vertical and volume regimes of the upper ocean water column (horizontal and temporal variability is implicitly assumed).

The SST variables defined here have been carefully considered by the GHRSST-PP Science Team in order to achieve the closest possible coincidence between what is defined and what can be measured operationally, bearing in mind current scientific knowledge and understanding of how the near surface thermal structure of the ocean behaves in nature. The following SST definitions are defined and explained according to the consensus reached at the 2nd (Donlon, 2002b) and 3rd GHRSST-PP workshops (Donlon et al., 2003b).

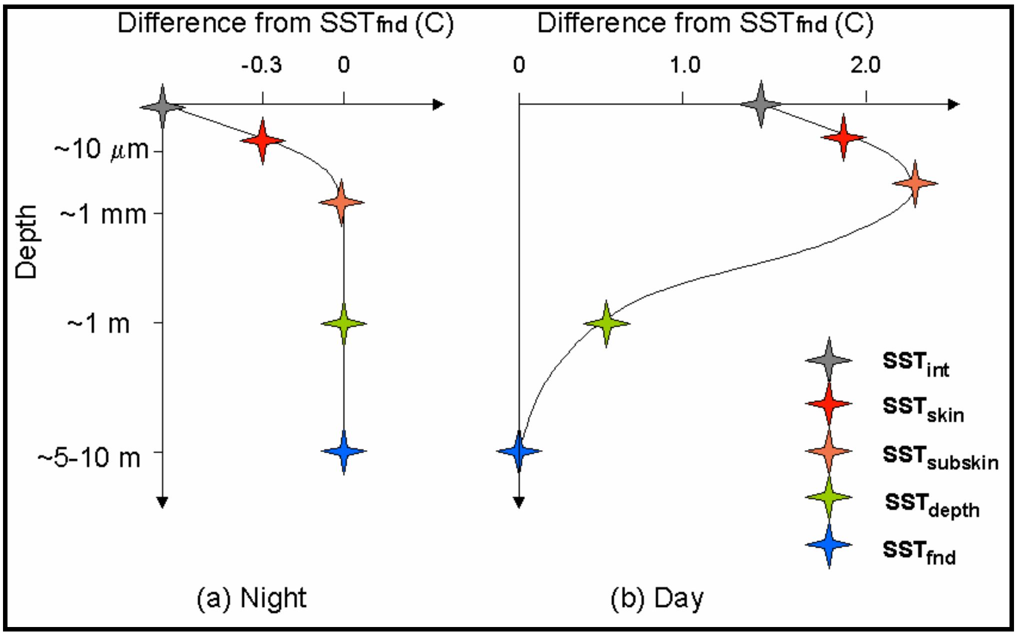


Figure 2.4.1 Schematic diagram showing (a) idealised night-time vertical temperature deviations from SSTfnd and (b) idealised day-time vertical temperature deviations from SSTfnd in the upper ocean.

2.4.1 The Foundation SST (SSTfnd)

The **foundation SST, SSTfnd**, is defined as the temperature of the water column immediately below the bottom of the diurnal thermocline, if there is one, or the SSTsubskin if there is not. It is named to indicate that it is the foundation temperature from which the growth of the diurnal thermocline develops each day. SSTfnd provides a connection with historical "bulk" SST measurements typically used as representative of the oceanic mixed layer temperature. This definition was adopted by GHRSST-PP at the Third GHRSST-PP Workshop (Donlon, 2003b) to provide a more precise, well-defined quantity than previous loosely defined "bulk" temperature quantities and consequently, a better representation of the mixed layer temperature. The SSTfnd product provides an SST that is free of any diurnal variations (daytime warming or nocturnal cooling). In general, SSTfnd will be similar to a nighttime minimum or pre-dawn value at depths of 1-5 m, but some differences could exist. Only in situ contact thermometry is able to measure SSTfnd. SSTfnd cannot be directly measured using either microwave or infrared satellite instruments. Analysis procedures must be used to estimate the SSTfnd from radiometric measurements of SSTskin and SSTsubskin.

2.4.2 The SST at depth (SSTdepth)

SSTdepth or SST(z) is the terminology adopted by GHRSST-PP to represent an *in situ* measurement near the surface of the ocean that is typically reported simply as SST or "bulk" SST. For example SST6m would refer to an SST measurement made at a depth of 6m. Without a clear statement of the precise depth at which the SST measurement was made, and the circumstances surrounding the measurement, such a sample lacks the information needed for comparison with, or validation of satellite-derived estimates of SST. The terminology has been introduced to encourage the reporting of depth (z) along with the temperature.

2.4.3 The subskin SST (SSTsubskin)

SSTsubskin represents the temperature at the base of the thermal skin layer. The difference between SSTint and SSTsubskin is related to the net flux of heat through the thermal skin layer. For practical purposes, SSTsubskin can be well approximated to the measurement of surface temperature by a microwave radiometer operating in the 6-11 GHz frequency range, but the relationship is neither direct nor invariant to changing physical conditions or to the specific geometry of the microwave measurements.

2.4.4 The Skin SST (SSTskin)

SSTskin is defined as the radiometric skin temperature measured by an infrared radiometer operating in the 10-12 µm spectral wave band. As such, it represents the actual temperature of the water at a depth of approximately 10-20 µm. This definition is chosen for consistency with the majority of satellite infrared measurements and ship mounted radiometer measurements available for *in situ* validation. SSTskin measurements are subject to both skin layer and diurnal warming effects and potentially exhibit a significant diurnal cycle.

2.4.5 The Interface SST (SSTint)

SSTint is a theoretical temperature at the precise air-sea interface. It represents the hypothetical temperature of the topmost layer of the water, but is of no practical use because it cannot be measured using current technology.

2.4.6 The diurnal cycle/variation of SST (DV)

For the GDSv1, diurnal cycles refers to changes in SST throughout a 24 hour period and thus includes warm stratified layers during the day and cool skin effects throughout a diurnal period, although the latter will be more pronounced at night.

2.4.7 List of acronyms

Table 2.4.7 Acronyms and abbreviations applicable to the GDS

AATSR	Advanced Along Track Scanning Radiometer
AMSR	Advanced Microwave Scanning Radiometer
APPW	Analysed Product Processing window
AUS	Applications and user services
AVHRR	Advanced Very High Resolution Radiometer
BT	Brightness temperature
Clw	Cloud liquid water content
CTD	Conductivity, temperature, depth (in situ ocean measurements)
DDS	Diagnostic data set
DODS	Distributed Oceanographic Data System
DV	Diurnal Variation
ECMWF	European Centre for Medium-range Weather Forecasting
EOEP	Earth Observation Envelope Programme (ESA)
ERRLOG	Operational Error log
ESA	European Space Agency
ESL	Expert support laboratory (ESA)
EURDAC	European GHRSST-PP RDAC coverage
FOAM	Forecasting ocean assimilation model
GDAC	Global data analysis centre
GDIP	GHRSST-PP development and implementation plan

GDS	In situ and satellite data integration processing model
GHRSST-PP	The GODAE High Resolution Sea Surface Temperature Pilot Project
GI	Global integration
GODAE	Global Ocean Data Assimilation Experiment
GOES	Geostationary operational environmental satellite
GOOS	Global ocean observing system
GTS	Global telemetry system
IODD	Input-output data definitions
IPCV	Infrared proximity confidence value
L2	Level-2
L2P	Level-2 data with added confidence flags after checking for gross errors, consistency and timelines. This family of data products provides the highest quality data obtained from a single sensor for a given processing window.
L3	Level-3
LAS	Live Access Server
MCSST	Multi-channel sea surface temperature
MDB	Match up database
MMR	Master Metadata Repository
MPCV	Microwave proximity confidence value
MSG	METEOSAT second Generation
MSG	Meteosat Second Generation
NASDA	National Space Development Agency of Japan
NCEP	National Center for Environmental Prediction (US)
NGSST	New Generation SST Project (Japan)
NOAA	National Ocean and Atmosphere Administration
NWP	Numerical Weather Prediction
OPLOG	Operational data log
PEMSA	Production of Enhanced Multi-sensor SST Analysis Project (U.S.)
PO.DAAC	Physical Oceanography Data Active Archive Centre (U.S.)
PW	Processing Window
RA-2	Radar Altimeter-2
RAN	Re-analysis
RD	Reference document (see section 1.5)
RDAC	Regional data assembly centre
RGTS	Regional and global task sharing
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SSI	Surface Solar Irradiance
SSM/I	Special sensor microwave imager
SSES	Single Sensor Error Statistics
SST	Sea Surface Temperature
TAG	Technical advisory group
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Mapping Mission
U	surface wind speed at 10m height
UHR	Ultra-high resolution
UIS	User information services
URD	User requirements document

VAL	Validation module of GHRSST-PP
WMO	World Meteorological Organisation
XBT	Expendable bathythermograph (In situ measurement of ocean temperature profiles)
XML	Extensible Markup Language

Overview of the GHRSST-PP Processing Specificationv1.0 (GDS-v1)

This section of the GDSv1 provides a general overview of the processing model introducing the data flow within the processor, the processor input and output data definitions (IODD), and the work packages (WP) that will be used to implement the processor. It may be used as a "map" of the GDS and identifies the interfaces between each specific WP that are expanded in technical detail within the subsequent sections and associated appendices.

3.1 Summary description of the GDSv1

Figure 3.1.1 shows a functional breakdown diagram of the GDSv1. It identifies the major components of the processing specification as work-packages (shown in light blue) together with their associated input and output data parameters (shown in red and green respectively). The GDS is designed to produce SST data products that satisfy the requirements of operational ocean forecast and prediction systems. The main requirements are:

1. An error estimate for each SST measurement including a bias and RMS.
2. Operational near real time data availability (within 6 hours of data acquisition)

The flow of data and information between each WP is indicated by the direction of connecting arrows. Each GDSv1 WP has been designed to encapsulate a distinct suite of activities, common to all RDAC or GDAC that has a definite input and output definition (IODD). In this way, the exchange and use of all data products within the GDSv1 is greatly simplified by referring to the WP interface IODD provided in section 3.4.

RDAC first ingest, in real time, regional coverage satellite and in situ L2 SST and auxiliary data streams from a variety of different data providers (WP-ID1 in Figure 3.1.1). Each L2 satellite SST data stream is quality controlled to check for gross errors, consistency and timelines. Data that are unacceptable for further use in the GDS are rejected.

Before SST data originating from different sources can be properly assimilated into ocean model systems or analysed together to provide new SST data products that capitalise on their synergy, a bias and rms. error estimate is required for each [pixel] measurement. Assigning error estimates to SST measurements is a fundamental requirement of the GDS. This is performed as a three stage process:

1. A quantitative "confidence value", having a value from 0 (no data) to 7 (highest confidence in the SST measurement), is derived for each pixel measurement based primarily on an estimate of the surface wind speed, proximity to other cloud contaminated/rain contaminated pixels and, proximity to a SST reference climatology.
2. Error estimates for each L2 SST measurement are based on the statistical analysis of a match-up database (MDB) containing L2 satellite SST data and near contemporaneous quality-controlled in situ SST observations (WP-ID3 in Figure 3.1.1). Single Sensor Error Statistics (SSES) are produced at regular intervals from an analysis of all MDB data records for a given sensor and correlated with sensor specific confidence values.

3. An appropriate SSES bias and RMS value is assigned to each L2 SST measurement based on the actual confidence value for that specific measurement and its associated SSES.

The format of data records held within the MDB is described in Appendix 4 and has been designed to be compatible with other SST MDB (*e.g.*, the Miami pathfinder MDB) in order to take advantage of these data and to further contribute to their development.

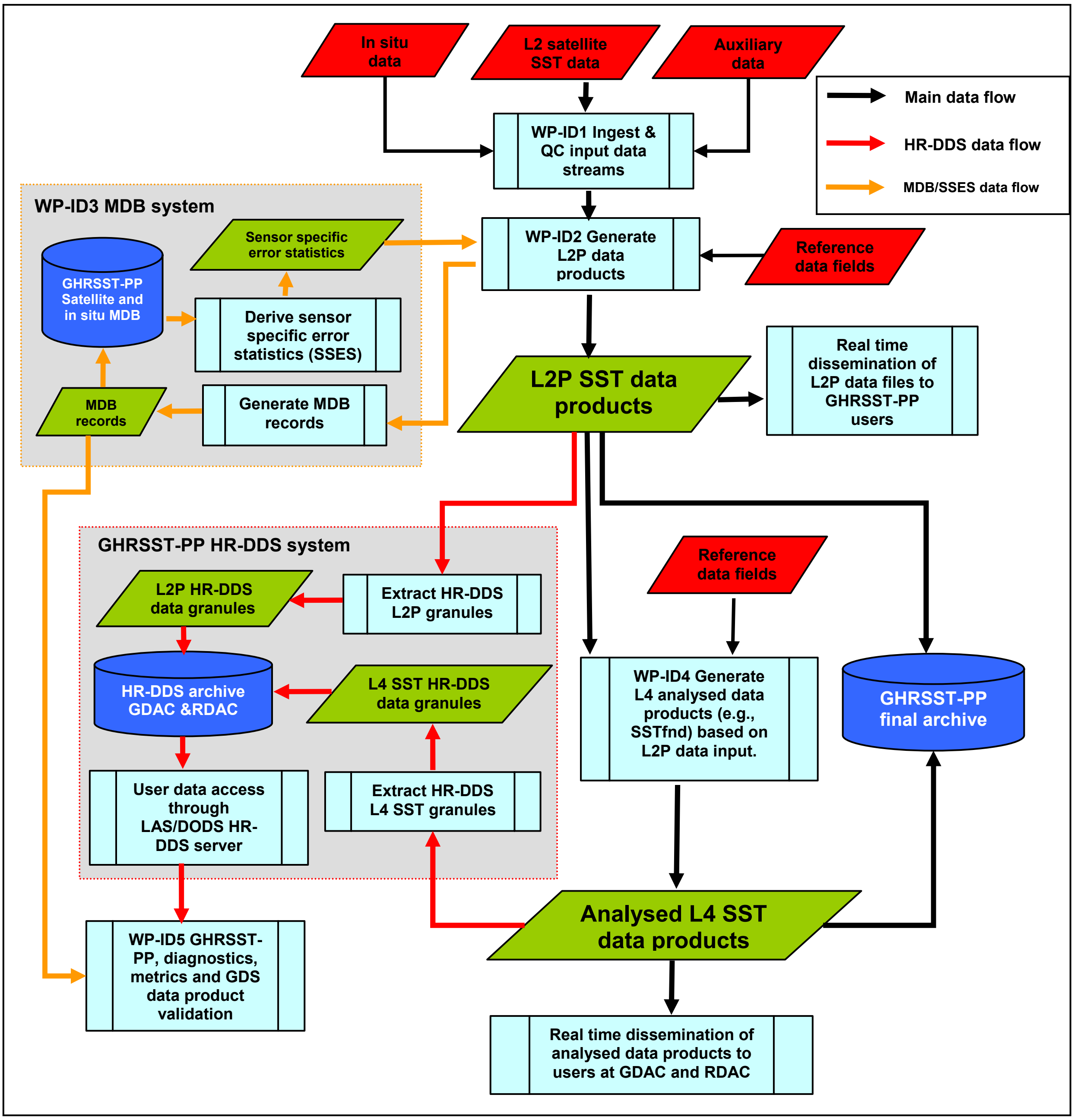


Figure 3.1.1 General overview of the GHRSST-PP GDSv1. Input data sources are shown in red, output parameters are shown in green and GDSv1 processing steps with a further breakdown are shown in light blue. Database storage is shown in dark blue. For each output parameter a corresponding metadata record is generated and delivered to the GHRSST-PP Master Metadata Repository (MMR) system. The MMR system has been omitted to preserve clarity in this figure.

Each L2 SST data file is then reformatted to a common GDSv1 L2 pre-processed data format (referred to as L2P and described in Appendix 1) that is designed to conform to the GODAE data sharing project (WP-ID2 in Figure 3.1.1). L2P data products are the lowest-level, common format data products produced by the GHRSST-PP and provide the "building blocks" for all other higher level data products. Each L2P data product contains L2 SST measurements together with confidence fields and SSES. L2P data products consist of the original L2 SST measurements (that are not re-gridded or modified) together with additional confidence and error estimates. A separate L2P data product is produced for each sensor at each RDAC. Operational users that have requested this type of SST data product may access L2P data products directly in real time from RDAC and GDAC. In addition, L2P data products provide a direct input to the GHRSST-PP Reanalysis project (Casey et al., 2003) and form the principal data archive for the GHRSST-PP.

GHRSST-PP high resolution diagnostic data set granules (HR-DDS, see Appendix A5) are extracted from L2P data products and are sent to the GHRSST-PP HR-DDS system for diagnostic analysis and validation of the GHRSST-PP GDS by the GHRSST-PP community. The HR-DDS provides a shared data resource that can be used to test, validate and refine the methods and data products that are produced by the GDS and is fully described in the HR-DDS implementation plan available at http://www.ghrsst-pp.org. Data access to the HR-DDS system will be through a Live Access Server (or variant) system. In addition, the HR-DDS provides a focus for commissioning the GDS and the production of on-going metrics required to assess the performance of the processor. The HR-DDS is therefore closely integrated within the GDS.

GHRSST-PP L2P data products will serve the requirements of many operational ocean modelling groups in real time. However, many users of GHRSST-PP SST data products (including operational modelling groups) request complete SST fields, free of gaps caused by clouds, rain or lack of data coverage, at regular intervals of between 6 and 24 hours. Some require an estimate of the sub-surface SST field that cannot be directly measured using current satellite techniques whereas others require an estimate of the surface skin temperature that is in direct contact with the overlying atmosphere and subject to considerable diurnal variability. To address these needs, L2P data will used in an optimal interpolation scheme (e.g., Reynolds and Smith, 1994; Reynolds et al, 2002; Guan and Kawamura, 2003, Murray et al. 2002, Murray et al. 1994, Fieguth et al., 1998;2000, Menemenelis et al. 1997) that is designed to

(a) Account for differences in spatial and temporal sampling characteristics of each data stream,
(b) Account for gaps in coverage due to the presence of cloud, rain or lack of data,
(c) Account for SST diurnal variability and retrieve an estimate of the subsurface temperature field (referred to as the foundation temperature, SSTfnd) and SST1m,
(d) Derive a measure of diurnal variability within the data product time domain to accompany the SSTfnd estimate and,
(e) Retrieve an estimate of the skin temperature of the ocean (SSTskin).

At 24 hour intervals, called Analysed Product Processing Windows (APPW) defined in Table 2.3.2, global coverage L2P data products are used in an analysis procedures that will generate an estimates of the subsurface SST (WP-ID6 in Figure 3.1.1). Clearly, it is extremely important that RDAC data products are provided promptly to GDAC in order that global coverage data products can be made available to the GHRSST-PP community in a timely manner. Two L4 data products are specified by the GDSv1:

1. The SST at depth of 1m (SST1m)
2. The foundation SST (SSTfnd) including an estimate of SSTskin diurnal variability (magnitude and phase).

Each L4 data product is formatted to a common GDSv1 L4 data format (Appendix A1) that contains the analysis at each grid cell together with quality control and error statistic information. L4 data products will be available in real time to the GHRSST-PP user community within 6 hours of an APPW. GHRSST-PP high resolution diagnostic data set granules (HR-DDS) are extracted from all L4 data products and are sent to the GHRSST-PP HR-DDS system for diagnostic analysis and MDB data records should be prepared and sent to the GHRSST-PP MDB system for product validation activities.

3.2 Input and output data definitions (IODD) for the GDSv1

Appendix A3 provides a complete description of the input data streams for the GDSv1 and the main outputs of the GDP are listed in Table 3.2.1.

Table 3.2.1 Summary of GDSv1 output

Product identifier	Descriptive name	Description	GDSv1 section	Data format definition
L2P-<filename>	L2 pre-processed data	Native L2 SST and auxiliary data that have been quality controlled and re-formatted to include confidence and error statistic data. <filename> is defined in Table A2.2	WP-ID2	Appendix A1.2 and A1.4
L4SST1m	GDAC L4 SST1m analysed data	Global coverage SST1m analysed data products for each APPW based on the New Generation SST method	WP-ID4	Appendix 1.2, A1.3 and A1.7
L4SSTfnd	GDAC L4 SSTfnd analysed data	Global coverage SSTfnd analysed data products for each APPW	WP-ID4	Appendix 1.2, A1.3 and A1.7
SSES	Sensor specific error statistics	Mean bias and rms. Error statistical relationships to sensor specific confidence values.	WP-ID3	N/A
MDB	Match up data base record	Near contemporaneous satellite and in situ data match up record.	WP-ID3	Appendix 4
L2P HR-DDS granules	L2P HR-DDS data granules	A high resolution diagnostic data set (HR-DDS) granule (2° x 2° latitude x longitude area) extracted from a L2P data product	Appendix A5	Appendix 1.2, A1.3 and A1.7
L4SST1m-HRDDS	L4SST1m high resolution DDS data granule	A high resolution diagnostic data set (HR-DDS) granule (2° x 2° latitude x longitude area) extracted from a L4SST1M data product	Appendix A5	Appendix 1.2, A1.3 and A1.7
L4SSTfnd-HRDDS	L4SSTfnd high resolution DDS data granule	A high resolution diagnostic data set (HR-DDS) granule (2° x 2° latitude x longitude area) extracted from a L4SSTfnd data product	Appendix A5	Appendix 1.2, A1.3 and A1.7

WP-ID1 Ingestion of Satellite and in situ data streams

WP-ID1	Ingestion of Satellite and in situ data streams
Work Package number:	**WP-ID1**
Leader:	**Ed Armstong (ed@seastar.jpl.nasa.gov)**
Aim	**To ingest real time satellite and in situ data streams from different data providers for use in the GHRSST-PP.**
1	**Objectives:** 1. Specify the Level-2 (L2) data streams that will be considered by the GDS. 2. Specify a data provider responsible for each L2 data stream. 3. Specify the primary and any secondary GDAD/RDAC entry points for each L2 input data stream. 4. Specify the limitations to the use of data arising from specific agreements pertaining to data use within the GHRSST-PP. 5. Specify the data transfer mechanism from data provider to GHRSST-PP primary entry point. 6. Prepare Master Metadata Repository (MMR) data set description (MMR-DSD) records for each L2 data stream. 7. Specify the procedures to register each L2 data set at the MMR. 8. Submit appropriate Level-2 Data Processing synopsis Report (L2_DPSR) to the GHRSST-PP OPLOG and ERRLOG logs to reflect the status of each data stream.
2	**Description:** This WP is dedicated to the operational ingestion of satellite and in situ data streams by GHRSST-PP GDAC and RDAC. The main purpose of this WP is to specify the data products and data exchange/provision mechanism that will be used within the GDSv1. In addition, MMR DSD metadata records should be prepared for each data stream at the MMR. Together with Appendix 3 this WP provides a reference document for L2 data streams used within the GDSv1.
3	**Inputs:** 1. L2 satellite data product definitions. 2. In situ observation definitions. 3. Data provider definitions and restrictions. 4. RDAC/GDAC data usage descriptions.
4	**Outputs:** 1. GDSv1 L2 data file description tables (Appendix 3). 2. Metadata received at MMR from RDAC and GDAC for each L2 data file ingested. 3. Entry into the OPLOG and/or ERRLOG processor log files. 4. E-mail to L2 data provider explaining L2 data ingestion problems. 5. MMR DSD records for each data stream. 6. L2 data received at GHRSST-PP primary entry point
5	**Acceptance tests:** 1. RDAC and GDAC MMR metadata formatted correctly for each GDS L2 data stream. Validated using the GDSv1 reference processor. 2. MMR metadata functional.

	3. Data provider points of contact established and table entry verified with point of contact. 4. HR-DDS established and DSR records ingested.
6	**Metric for performance assessment:** 1. MMR metadata records provided and ingested at MMR in real time. MMR rejection rate < 2%. 2. Data provider promptly notified in case of problems with L2 data stream. 3. OPLOG and/or ERRLOG log entries are timely. 4. Newsletter provided to data providers informing of data usage statistics based on MMR records.

Conventional L2 SST, auxiliary satellite (e.g., wind speed and surface solar irradiance (SSI)) and in situ data streams are used by the GDSv1. The GDSv1 does not consider L0, L1a or L1B data streams although the latter may be useful for certain HR-DDS sites for diagnostic analysis (see WP-ID5). Each L2 data stream will be formatted according to a well specified format by a specific data provider. Note that L2 data streams from the same sensor may have different formats depending on the data provider. In general, L2 data are ingested at RDAC and in some cases at GDAC centres. Table 4.1 describes the L2 data sets that are available to the GHRSST-PP during the period 2003-2007 categorised by data type.

Table 4.1 Real Time global and regional satellite and in situ SST data streams available to the GHRSST-PP during the period 2003-2007. In situ data streams are those providing data within the upper 8 m of the water column. Acronym definitions are given in Appendix 2 Table A2.2.

Data type	Global coverage	Regional coverage
Operational satellite data streams	AVHRR16-G AVHRR17-G AMSR AMSRE	AVHRR16-L AVHRR17-L ATS_MET_2P[$] GMS-VISSR[#] SEVIRI[#] GOESE[#] METEOSAT5[#] GOESW[#]
Research satellite data streams	GLI (no NRT)[#] MODIS-A (no NRT)[#] MODIS-T (no NRT)[#]	ATS_NR__2P[#] TMI[$]
Operational in situ data streams	None	ARGO SOOP VOSCL DRI TAO PIRATA JMA ODAS
Research in situ data streams	None	RADIOM CRUISE

[#] High data rate [$] Low data rate

Data streams that have no real time capability will be used in GHRSST-PP reanalysis products and within the GHRSST-PP High Resolution Diagnostic Data Set (HR-DDS) in a delayed mode.

Figure 4.1 provides a functional breakdown diagram that identifies the main data processing tasks within WP-ID1. These are:

1. Establish real time operational access to L2 data streams at RDAC and GDAC centres e (i.e. regional or global).
2. Evaluate each L2 data file for integrity when ingested at an RDAC or GDAC facility.
3. Determine if the L2 data file should be accepted into the GDSv1. If not, then raise an error report and send this to both the ERRLOG log and to the data provider.
4. Admit the L2 data stream into the GDSv1 by preparing a Metadata Master Repository (MMR) File Record (MMR-FR) metadata record and submitting this to the MMR system. A MMR-DSD record must be created at the MMR for each L2 data stream described in Appendix 3.
5. Make an entry into the OPLOG processor log describing the activity and the activity status.

Each sub task is described in the following sections.

WP-ID1.1. Access to Level-2 input data streams

Appendix 3 describes the satellite and in situ data streams that will be considered by the GDS-v1. Appendix A3.1 defines the satellite L2 SST data streams, Appendix A3.2 the L2 auxiliary satellite data streams and Appendix A3.3 the L2 in situ data streams. A unique identification code for each L2 data stream is given that will be referred to throughout the GDS-v1. Also specified are the general characteristics of each data stream including the data provider, data delivery mechanism, any data agreement that is applicable together with the primary and secondary RDAC/GDAC entry points. All of this information will be used to generate MMR-DSD records for each L2 data stream that is used by the GDSv1 according to the specifications provided in Appendix A6 Table A6.2.1.

The data streams within the GDSv1 are dynamic in the sense that new data sets will become available, certain data sets may take priority in given circumstances (e.g., aerosol contamination) and regional preferences based on the ease of data access will exist at RDAC.

WP-ID1.2 Evaluate each L2 data file

When a L2 a file enters a GDAC or RDAC processing facility it typically resides on a temporary staging area at an RDAC or GDAC centre following transfer from the data provider. In some cases, the delivery of data may be incomplete because for example, data connections could be lost, corruption of the data may have occurred during delivery, or the original file was incorrectly generated. The purpose of WP-ID1.1 processing step is to verify three concerns:

(a) Was data transfer successful and is the data file complete?,
(b) Is the data file correctly formatted according to the specifications provided by the data provider?
(c) Was the delivery timely and are the L2 data eligible for use in the current PW?

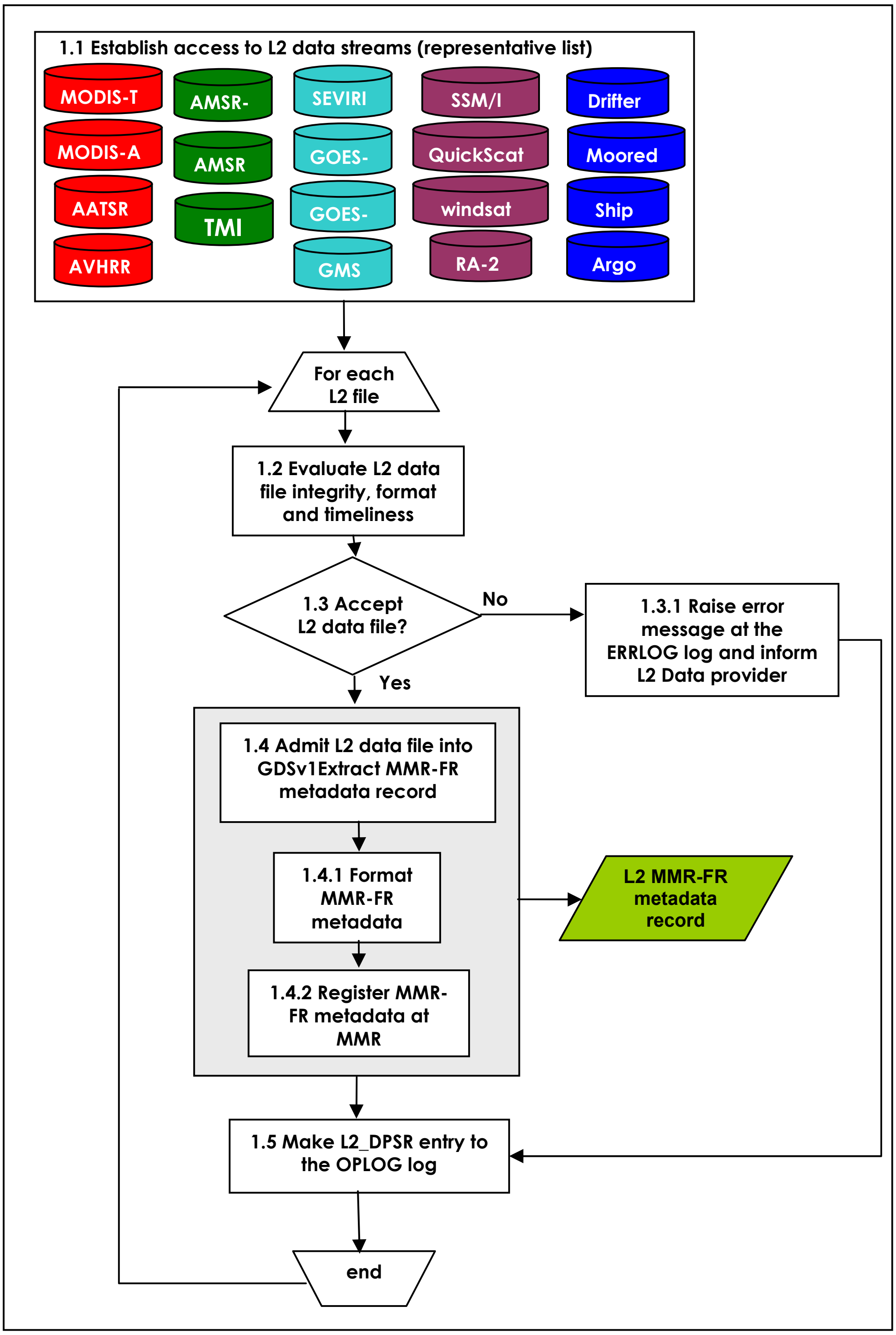


Figure 4.1.1 Functional breakdown of GDSv1 WP-ID1 identifying all sub tasks.

Individual RDAC and GDAC are responsible for implementing local procedures that are able to assess the integrity, format, and delivery timeliness of each L2 data file. The aim is to provide the RDAC/GDAC processing system with an immediate indication of a data delivery problem. All procedures should focus on reliability and simplicity rather than schemes that are likely to introduce unnecessary complexity into the data processing system.

WP-ID1.3 Admit L2 data file into the GDSv1 system

Based on the assessment made by RDAC in WP-ID1.21 a decision is made to either accept a L2 data stream or to reject the data stream from the GDSv1. The GDSv1 specifies the following rules for this purpose:

- **Rule 1.3.1:** A L2 data file should be admitted into the GDSvv1 processing system by an RDAC or GDAC only if the data file has been verified as complete, correctly formatted and that the delivery was timely with respect to the current PW.

The GDSv1 recommends that in order to identify a L2 data file that has been admitted into the GDSv1 system by an RDAC or GDAC, the data file should be appended to the data set code according to the following specification:

<processing centre code>-<data set code>-L2-<original L2 filename>

e.g. **`USA-AVHRR16_G-L2-12052899898834AVH_GAC.l1b`**

where <processing centre code> is defined in Table A2.1 and <data set code> is defined in Table A2.2. Renaming a L2 file in this way provides a simple method to identify data that a L2 data file has passed the initial data delivery/integrity/timeliness tests from those that have not while preserving the original L2 filename. This is important in order that data files may be traced back to data providers and RDAC/GDAC centres if problems are found at a higher level within the GDSv1 processing chain.

WP-ID1.3.1 Inform data provider of L2 data delivery problem

If a L2 data file has been rejected by an RDAC/GDAC, an error must be raised and logged at the ERRLOG system as described in Appendix A7.2 using the appropriate ERRLOG code and message format. In addition, an e-mail message should be sent to the L2 data provider that explains that the L2 data has been rejected form the GHRSST-PP GDS stating the reason why. A copy of this e-mail should be sent to the ERRLOG system log. In addition, an entry into the OPLOG log should be made as described in WP-ID1.5.

WP-ID1.4 Create and register an MMR-FR for each L2 data stream

In order for L2 data files to become "visible" to the global GDSv1 system and all other RDAC and GDAC processing centres, a MMR-FR metadata record must be prepared and successfully submitted to the GDS MMR. For each L2 data file that is admitted into the GDSv1, a MMR-FR must be prepared and registered at the MMR by the responsible RDAC/GDAC processor. Once a MMR-FR is successfully registered at the MMR system, all GHRSST-PP processing centres are aware of which L2 data files have entered the GDSv1 system. Only after registration at the MMR is a L2 data set considered to have entered the GDSv1. Note that **no** MMR-FR metadata are prepared for rejected L2 data streams.

WP-ID1.4.1 Format L2 MMR-FR metadata record

A MMR-FR metadata record should be prepared for each L2 data file according to the specification provided in Appendix A6.2.

WP-ID1.4.2 Register L2 MMR-FR metadata record with MMR system

MMR-FR metadata records should be delivered to the MMR system as soon as possible after the L2 data set has been evaluated following the procedure described in Appendix A6.4.

WP-ID1.5 Send an entry to the OPLOG log

In order to maintain a record of L2 data problems, a L2 data processing synopsis report (L2_DPSR) entry within the GDS operational log should be made via e-mail as explained in Appendix 7 indicating if each L2 data stream has been accepted or rejected from the GDSv1.

WP-ID2: Generation of GHRSST-PP Level-2 Pre-processed Data files (L2P).

WP-ID2	Generation of GHRSST-PP Level-2 Pre-processed Data (L2P)
Work Package number :	**WP-ID2**
Leaders:	**Craig Donlon (craig.donlon@jrc.it) Richard Reynolds (Richard.W.Reynolds@noaa.gov>**
Aim	**To specify operational methods for the quality control of GHRSST-PP input data streams specified in WP-ID1 and generate L2P data files**
1	**Objectives:** 1. Extract confidence data provided with L2 input data. 2. Quality control L2 input data and derive pixel confidence data. 3. Determine if L2 data are suitable for further use in the GDSv1. If not inform the data provider why L2 data have not been used in the GDSv1. 4. Prepare and output L2P data files and associated Master Metadata Repository File Records (MMR-FR) 5. Submit appropriate L2P Data Processing synopsis Reports (L2P_DPSR) to the OPLOG and ERRLOG logs to reflect the status of each data file.
2	**Description:** This WP is focussed on the pre processing and quality control of L2 data streams at RDAC and GDAC. The main output is L2P data files which consist of the native L2 observations together with additional pixel confidence data stored in GDSv1 netCDF format. L2P data products form the basic "building blocks" for all other GHRSST-PP data products and are a necessary precursor to their use in more complex analysis procedures (e.g., Reynolds and Smith, 1994). L2P data files will be made available to the GHRSST-PP operational community in real time. Each input data file is assessed, pixel by pixel, using gross error checking rules. A configuration file is used to store all QC parameter limits and thresholds that will be revised as required. Confidence data are extracted from L2 native data files if present or they are derived using reference data and sensor specific error statistics (SSES, generated by WP-ID3). If L2 data are unsuitable for further use in the GDSv1 then a message is sent to the L2 data provider and to the ERRLOG log informing why L2 data have not been used. If L2 data are suitable for further use within the GDS then a L2P data file is created and a L2P MMR-FR created for each L2P data file.
3	**Inputs:** 1. L2 data files 2. Reference data sets 3. GDSv1 L2P configuration file 4. Error checking rules
4	**Outputs:** 1. GHRSST-PP L2P data files 2. Rejection/acceptance message to data provider and ERRLOG log 3. L2P MMR-FR metadata record 4. OPLOG L2P_DPSR

5	**Acceptance tests:** 1. L2P output data products based on the GDSv1 test input data set produced at each RDAC and GDAC should be identical to those produced by the GDSv1 reference data processor. 2. L2P measurement data and confidence data records are identical to actual L2 confidence data 3. HR-DDS interface is functional 4. Data provider messages are sent and received correctly 5. Gross error checking procedures successfully flag unsuitable data and admit useful data with minimal data loss 6. MMR-FR metadata entries are timely and accurate
6	**Metric for performance assessment:** 1. L2P data are produced in a timely manner 2. L2P MMR-FR metadata records provided and ingested at MMR in real time. MMR rejection rate < 2% 3. L2 data provider promptly notified in case of problems with L2 data stream 4. OPLOG and/or ERRLOG log entries are timely 5. Delay between RDAC L2 data file ingestion and production of L2P data file should not affect the production of higher-order (L4*) data files i.e. PW and APPW should be used as constraints.

The purpose of this WP is to derive Level-2 Pre processed (L2P) GHRSST-PP data files by testing each L2 SST measurement using quantitative quality control tests, deriving pixel confidence data and assigning error estimates. In this way, problematic L2 data can be identified at the start of the GDSv1 data processing effort leaving as much time as possible to correct faults. Figure 5.1 shows the functional breakdown diagram for WP-ID2.

L2 input data sets ingested at RDAC and GDAC (WP-ID1) together with reference data fields (described in Appendix A3.1) and are input into a pre-processing and quality control procedure. Reference data and statistical tests are used to assign a confidence value to each pixel of the input data stream (WP-ID2.1). Based on the results of WP-ID2.1, single sensor error statistic (SSES) bias and RMS error values are assigned to each L2 measurement (explained in WP-ID3).

If there are errors or other problems with L2 data files (e.g., overwhelming cloud contamination, no data values, or values are outside of expected ranges) an e-mail is sent to the L2 data provider informing them that the L2 data have not been used in the GDS and a brief synopsis of why the data have been rejected is provided (WP-ID2.2.1). An appropriate entry into the ERRLOG log is made in this circumstance.

L2P data files are formatted as GDSv1 netCDF files (WP-ID1.2.3) as described in Appendix 1. For each L2P data file a MMR-FR metadata record (Appendix 6) is generated and registered at the MMR (WP-ID2.4). Finally, an L2P Data Processing synopsis Report (L2P_DPSR) entry is made in the OPLOG log file (Appendix A7.1) summarising the conversion of L2 data files to L2P data files. The difference between L2 and L2P data files is that additional confidence data and sensor specific error estimates for each pixel value have been assigned and stored with the L2 SST measurements. No modification to the L2 data values have been made during the conversion process. L2P data products will be available to the operational user community in real time.

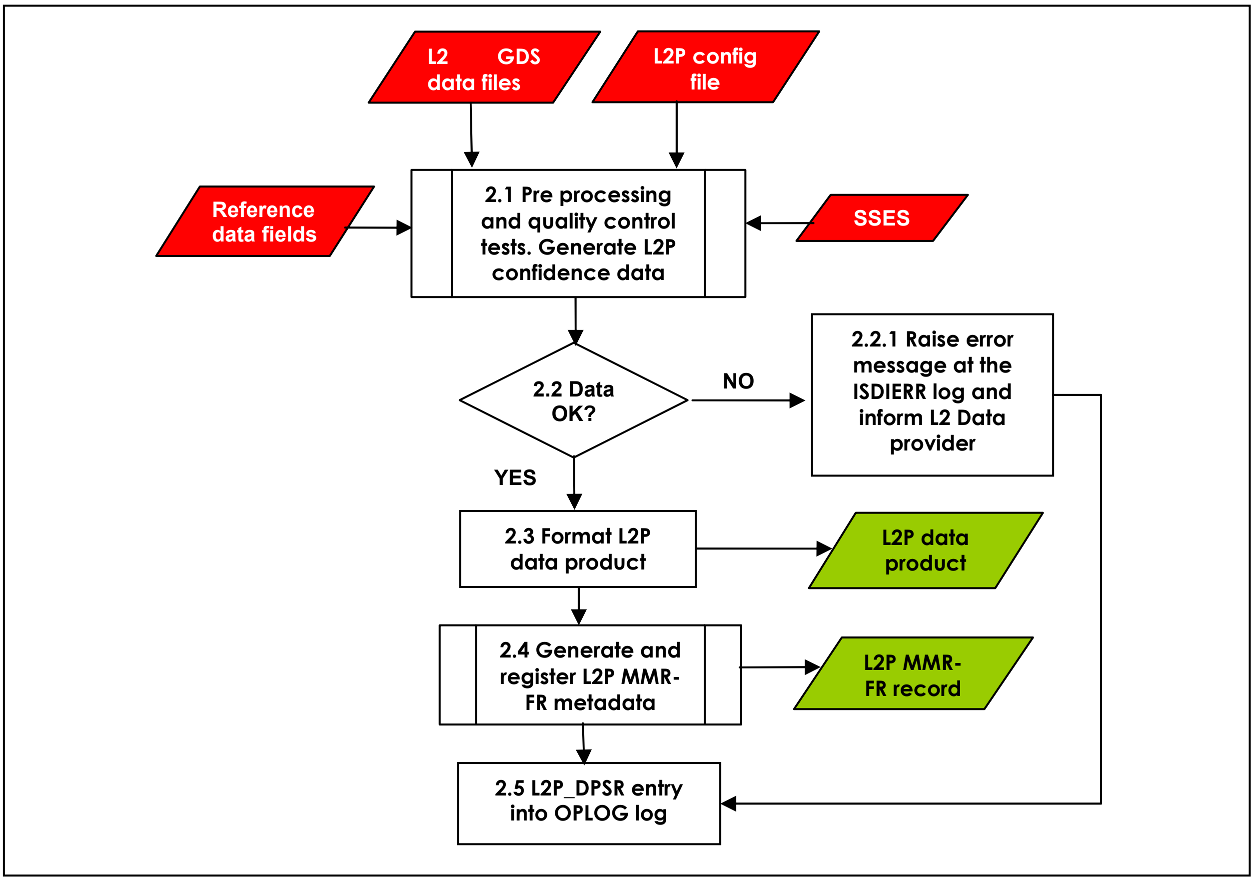


Figure 5.1 Functional breakdown of WP-ID2 identifying each major WP sub task.

WP-ID2.1 Quality control of L2 satellite data sets and assignment of pixel confidence data.

L2P confidence data are derived for every measurement within a L2 data stream and include parameters and coded flags to specify various characteristics of the L2 data stream (e.g., cloud contamination cosmetic pixel values etc.). Pixel confidence data is carried through the entire GDS-v1 chain.

As there are a number of different L2 data streams that the GDSv1 must consider (SST, wind speed, solar radiation etc.) several different quality control (QC) test must be specified according to each particular data stream. The specific thresholds and limits that are used to QC each L2 data file are stored in a system wide L2P configuration file. This allows for modifications to threshold values and other configurable parameters without the need for any software code changes.

The following sections describe the QC tests and pre-processing of L2 data files that should be performed on each generic L2 data file. Special emphasis is given to SST data streams which are the focus of the GDS processor in WP-ID2.1.1. Auxiliary L2 streams (including wind speed and surface solar irradiance (SSI)) are only used as indicators within the GDSv1 and it is left to the RDAC and GDAC to determine the level of QC that is required to use these data with confidence. Only simple tests are specified for wind speed (WP-ID2.1.2) and SSI (WP-ID2.1.3) data streams in the GDS.

WP-ID2.1.1 SST quality control tests

The purpose of these simple tests is to establish any obvious problems that suggest that a given L2 measurement is

- erroneous or,
- of limited further use within the GDS (cloud/rain contaminated, a cosmetic pixel, out of specified range, etc.)

Figure 5.1.1.1 shows a functional breakdown diagram of WP-ID2.1 that identifies the pre processing and confidence data extraction processing steps for L2 satellite SST data files. The output of these processing steps is a set of confidence data, SSES and an estimate of diurnal excursion for each pixel. Table A1.4.2 in Appendix 1 describes the format of GDSv1 L2P pixel confidence data record. Each processing step is described in detail in the following sections.

WP-ID2.1.1.1 No L2 data flag test

This test is used to set the L2P confidence data record variable **`no_L2_data`** defined in Table A1.4.2 in Appendix 1. L2 data streams should provide a flag or a data value indicating that no data are available for this particular pixel. This should be used to set the **`no_L2_data`** flag of the GDSv1 L2 confidence record. The GDSv1 specifies the following rule:

- **Rule 2.1.1.1:** If no L2 pixel SST data are available, determined by the L2 pixel value or an associated L2 confidence flag, the **`no_L2_data`** flag of the L2P confidence data record for that pixel must be set.

WP-ID2.1.1.2 Gross SST limit check

This test is used to set the L2P confidence data record variables **`SST_Bad`** and **`ΔTmin`** defined in Table A1.4.2 in Appendix 1. This test identifies data that fall outside of acceptable range of global SST values and identifies unrealistic SST values. The GDSv1 specifies the following rule:

- **Rule 2.1.1.2a:** A valid SST observation must lie within a temperature range of **`LowSSTLimit`** < SST < **`HighSSTLimit`**. If the pixel SST value is out of these limits, the **`SST_Bad`** flag must be set in the L2P confidence record for that pixel.
- **Rule 2.1.1.2b:** The difference between the pixel SST value (SSTobs) and the reference SST field[1], SSTref (defined in Appendix 3, Table A3.1.1), **`ΔTmin`** is derived using:

$$\Delta \mathrm{Tmin} = \mathrm{SSTref} - \mathrm{SSTobs} \qquad \text{(Eq. 2.1.1.2)}$$

The ΔTmin value should be inserted into the **`ΔTmin`** field of the L2P confidence data record for the pixel in question.

[1] This reference SST is SST1m derived as a 10 day "coldest" climatology from Pathfinder SST data products (Faugere et al. (2001)). SST values for a given location and season that are significantly beneath this climatological value indicate the presence of cloud contamination.

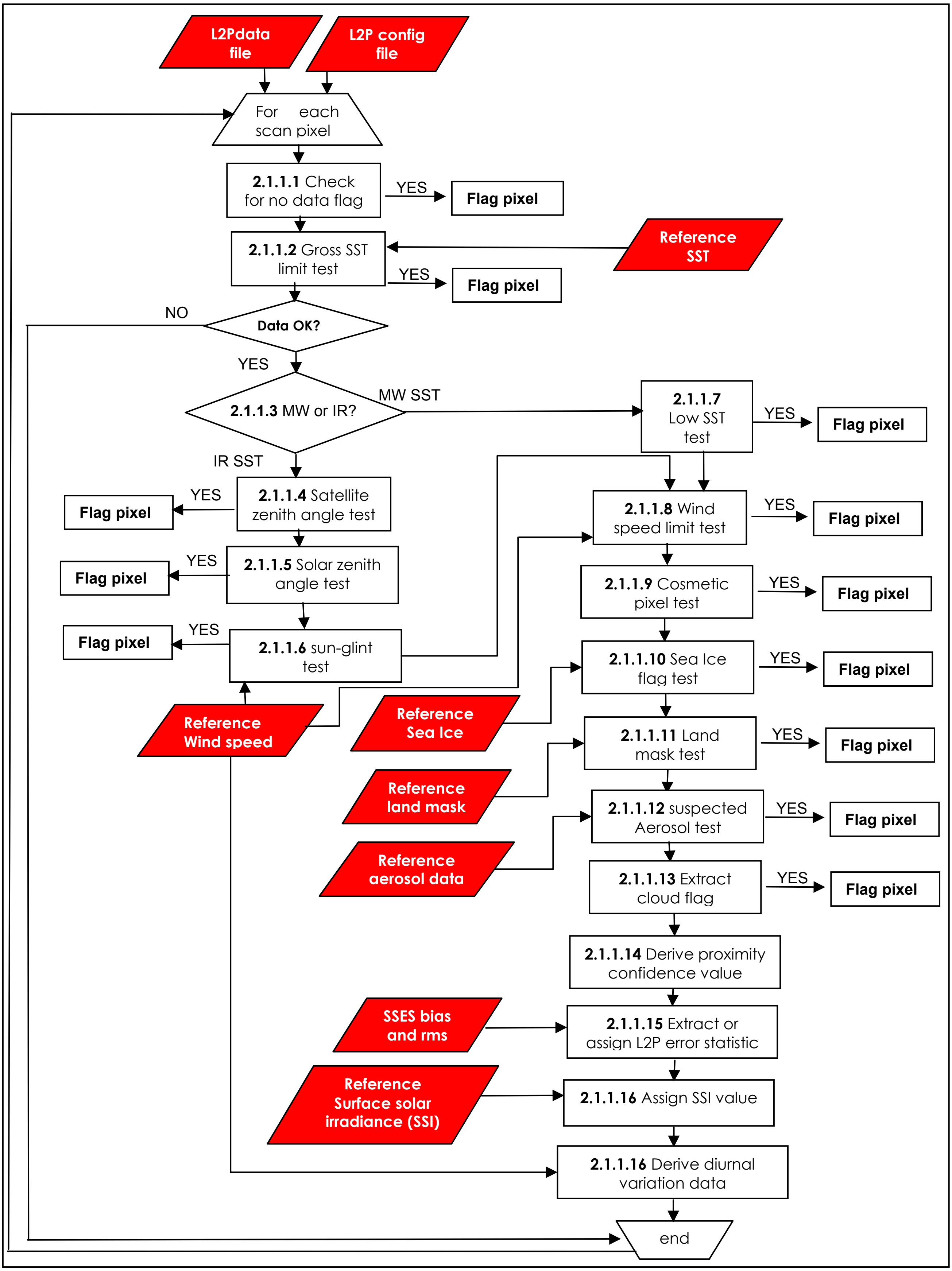


Figure 5.1.1.1 Functional breakdown diagram of pixel-by-pixel production of L2P data products. A L2P configuration file is used to store appropriate thresholds and reference values.

WP-ID2.1.1.3 Determination of data type (Microwave or Infrared)
A number of QC tests are specific to infrared SST data streams. The data processing bifurcates based on the type of L2 data stream which may be either microwave SST or infrared SST.

WP-ID2.1.1.4 Satellite zenith angle (satZA) test
This test is used to set the L2P confidence data record variable **`satZA`** and **`satZA_bad`** defined in Table A1.4.2 in Appendix 1. At high satellite zenith angles, the longer atmospheric path between the satellite sensor and ocean surface may significantly attenuate the water leaving signal beyond the specification of the SST algorithms used to derive the SST value. For infrared satellite sensors, the satellite zenith angle (satZA) should not be too large.
Many L2 data streams provide the satellite zenith angle as part of their confidence/ancillary data and this value may be used to set the **`satZA`** field of the L2P confidence data record (Table A1.4.2 in Appendix 1) if available. The GDSv1 specifies the following rules:

- **Rule 2.1.1.4a:** If a L2 satZA value exists, this value should be used to set the **`satZA`** field of the L2P confidence data record for that pixel.
- **Rule 2.1.1.4b:** If no L2 satZA value exists for a pixel measurement, an appropriate value should be calculated and used to set the **`satZA`** field of the L2P confidence data record for that pixel.
- **Rule 2.1.1.4c:** For all polar orbiting infrared satellite data products if the satZA > **`POsatZA_Limit`** threshold as defined in the L2Pconfiguration file, the pixel should be flagged using the **`satZA_bad`** flag of the L2P confidence record.
- **Rule 2.1.1.4d:** For all geostationary orbit infrared satellite L2 data products where **`satZA`** > **`GEOsatZA_Limit`** threshold as defined in the L2Pconfiguration file, the pixel should be flagged using the **`satZA_bad`** flag of the L2P confidence record.

WP-ID2.1.1.5 Solar zenith angle (solZA) test
This test is used to set the L2P confidence data record variable **`solZA`** defined in Table A1.4.2 in Appendix 1. Many L2 data streams provide the **`solZA`** as part of their confidence/ancillary data and this value may be used for the L2P confidence value of **`solZA`** if available. The GDSv1 specifies the following rules:

- **Rule 2.1.1.5a** If a L2 solZA value exists, this should be used to set the L2P confidence data record **`solZA`** value for that pixel.
- **Rule 2.1.1.5b** If no L2 solZA value exists, this should be calculated and used to set the L2P confidence data record **`solZA`** value for that pixel.

WP-ID2.1.1.6 Sun glint test
This test is used to set the L2P confidence data record variable **`SunGlint`** defined in Table A1.4.2 in Appendix 1. Depending on the local wind speed and sensor-sun view-geometry, sun-glint areas may be present in L2 data streams. Normal sun glint occurs in reflection from the oceans when the sun is behind and to the side of the satellite. Sun glint may degrade the SST value retrieved by affecting the quality of the cloud mask used by an infrared radiometer and consequently these regions should be flagged accordingly. Note that for polar orbiting satellite systems the problem is quite limited, but for geosynchronous orbit sensors it is unavoidable.

The GDSv1 specifies the following rules:

- **Rule 2.1.1.6a:** If an indication of sun-glint is provided with a L2 pixel value, this should be used to set the **`SunGlint`** flag of the corresponding L2P confidence data record.
- **Rule 2.1.1.6b:** If an indication of sun-glint is not provided with a L2 pixel value, an appropriate method should be used to set the **`SunGlint`** flag of the L2P confidence data record that is based on the geometry of sun and satellite position and surface roughness (e.g., Gardashov and Barla, 2001).

WP-ID2.1.1.7 Low value SST test for Microwave SST

The TRMM TMI SST data are suspect at lower SST values due to the reduced sensitivity of the TMI radiometer. At temperatures below 12ºC, TMI SST shows a warm bias relative to Reynolds OIv2 which is thought to be related to the reduced sensitivity of the 10.6 GHz TMI channel (used for SST retrieval) at these temperatures. This test is used to set the L2P confidence data record variable **`MW_SST_flag`** defined in Table A1.4.2 in Appendix 1. The GDSv1 specifies the following rule:

- **Rule 2.1.1.7:** If a TRMM TMI SST value is less than **`LowTMISSTLimit`** the measurement must be flagged in the corresponding L2P pixel confidence record using the **`MW_SST_flag`** field which should be set according to Table 2.1.1.8.

WP-ID2.1.1.8 Wind speed limit tests

This test is used to set the L2P confidence data record variables **`Wspd`** and **`MW_SST_flag`** defined in Table A1.4.2 in Appendix 1. A near contemporaneous wind speed estimate[2] is required for assigning a diurnal signal confidence value and as an indicator of the turbulent state of the air sea interface. At this data processing step a near contemporaneous wind speed value for each L2 pixel is coded and stored in the associated L2P confidence record. The GDSv1 specifies the following rules:

- **Rule 2.1.1.8a:** A 10m surface wind speed value, coded as described in Table A1.4.2, should be assigned to each L2 pixel using the GDSv1 L2P confidence data variable **`Wspd`**. Where possible, wind speed values should be extracted from available L2P wind speed measurements that are closest to the L2 SST measurement time in both space and in time.
- **Rule 2.1.1.8b:** In the case of microwave SST measurements, the use of associated microwave 10m wind speed measurements obtained from the same instrument providing the SST measurement may be used to set the L2P confidence data variable **`Wspd`**.
- **Rule 2.1.1.8c:** In the absence of a surface wind speed measurement, a NWP estimated 10m surface wind speed (See Appendix A3, Table A3.1.1) should be used to set the L2P confidence data variable **`Wspd`**. The value should be indicated as a negative number to indicate that it is not a measured value.

Above a wind speed of ~10ms^{-1}, there is an increase in the standard deviation of MW SST data due to surface foam and roughness effects. The code values shown in Table 2.1.1.8 should be used to set the L2P confidence data field **`MW_SST_flag`** data appropriately based on Rules 2.1.1.8d-e.

---

[2] An NWP field may be used here.

Table 2.1.1.8 L2P MW SST code values for `MW_SST_flag` field. The lowest value takes precedence in situations where two criteria are fulfilled.

Code	Description
0	SST value acceptable
1	Failed high wind speed test
2	Failed low wind speed test
3	TMI Failed low SST test

- **Rule 2.1.1.8d:** If the wind speed is less than **`MWLowWindThreshold`** as defined in the L2P configuration file, the corresponding **`MW_SST_flag`** field of the L2P confidence data record should be set according to the value shown in Table 2.1.1.8.
- **Rule 2.1.1.8e:** If the wind speed is less than **`MWHighWindThreshold`** as defined in the L2P configuration file, the corresponding **`MW_SST_flag`** field of the L2P confidence data record should be set according to the value shown in Table 2.1.1.8.

WP-ID2.1.1.9 Cosmetic data test

This test is used to set the L2P confidence data record variable **`CosmeticFill`** defined in Table A1.4.2 in Appendix 1. Some L2 data streams contain pixel values that are cosmetic rather than measured in order to provide a complete data product[3]. The purpose of this test is to extract the status of a L2 cosmetic pixel value flag (if such a flag is present in the native L2 data stream). The following GDSv1 rule is specified:

- **Rule 2.1.1.9:** If a L2 data stream includes information that indicates that a pixel value is cosmetic rather than measured, the L2P confidence record flag **`CosmeticFill`** for that pixel should be set accordingly. If no such flag is present the GDSv1 assumes the pixel value to be a measurement.

WP-ID2.1.1.10 Sea ice Flag test

This test is used to set the L2P confidence data record variable **`FractionalSeaIce`** defined in Table A1.4.2 in Appendix 1. Some L2 data streams provide a sea ice flag to indicate that the SST measurement is contaminated in part or wholly by sea ice. In cases where no such data are present in the L2 data stream a reference sea ice data product should be used to assess the likelihood of sea ice contamination. The GDSv1 specifies the following rules:

- **Rule 2.1.1.10a** If a L2 pixel sea ice flag exists, this should be used to set the **`FractionalSeaIce`** flag of the L2P confidence data record according to the code values defined in Table A1.4.2 in Appendix 1.

  Rule 2.1.1.10b If a L2 pixel sea ice flag does not exist, then a sea ice flag derived from the reference sea ice data set defined in Appendix A3.1 should be used to determine the value of the L2P confidence flag **`FractionalSeaIce`**.
- **Rule 2.1.1.10c** If no sea ice values are available from the reference data set, then a sea ice flag derived from AMSRE or AMSR gridded SST data product that is nearest in both space and time should be used to set the value of the L2P confidence flag **`FractionalSeaIce`**.

[3] For example, the ENVISAT ATS_NR__2P data stream may contain cosmetically filled pixels. These data are used to fill in data gaps across the swath following rectification of the curved ground swath made by the conical scan geometry used by the instrument.

WP-ID2.1.1.11 Land mask flag test

Some L2 data streams provide a sea land flag as part of their confidence/ancillary data and this value may be used to set the value of the L2P confidence data variable **Land_sea_flag**. The GDSv1 specifies the following rule:

- **Rule 2.1.1.11** If a native L2 pixel land flag exists, this should be used to set **Land_sea_flag** in the corresponding L2P confidence data record using the code value defined in Table 2.1.1.11.
- **Rule 5.1.1.12** If a native L2 pixel land flag does not exist, then the coded land flag extracted from the reference land mask (see Table 2.1.1.11) should be used to set **Land_sea_flag** in the corresponding L2P confidence data record.

Table 2.1.1.11 MODIS/Terra Land Cover Type 96-Day L3 Global 1km ISIN Grid Land water mask coded values. See Appendix A3 Table A3.1.1 for a full reference.

Value	Description
0	Shallow ocean
1	Land (Nothing else but land)
2	Ocean coastlines and lake
3	Shallow inland water
4	Ephemeral water
5	Deep inland water
6	Moderate or continental ocean
7	Deep ocean

WP-ID2.1.1.12 Suspected Aerosol presence flag test.

This test is used to set the L2P confidence data record variable **Aerosol_present** defined in Table A1.4.2 in Appendix 1. The GDSv1 specifies the following rules:

- **Rule 2.1.1.12a:** If an indication of aerosol contamination is provided in a native L2 data stream for a given pixel, this should be used to set the corresponding L2P confidence data record field **Aerosol_present**.
- **Rule 5.1.1.14:** If an indication of aerosol contamination is not provided in a native L2 data stream for a given pixel then an aerosol flag derived from the reference aerosol data set (defined in Appendix A3.1) that is nearest to the SST measurement in space and in time should be used to set the corresponding L2P confidence data record field **Aerosol_present**.

WP-ID2.1.1.13 Cloud flag test

This test is used to set the L2P confidence data record variable **Cloud** defined in Table A1.4.2 in Appendix 1. L2 data streams should provide a flag or a data value indicating that the L2 pixel was flagged as cloudy. The GDSv1 specifies the following rule:

- **Rule 2.1.1.13:** If L2 pixel data values or associated confidence data indicate a given pixel to be classed as cloudy, the **Cloud** flag of the SDI-PMv1 L2P confidence data record must be set.

WP-ID2.1.1.14 Derivation of L2P confidence value

This procedure is used to set the L2P confidence data record variable **Proximity_Confidence** defined in Table A1.4.2 in Appendix 1. A proximity confidence value is specified for each pixel based on the "proximity" of that pixel to several different criteria known to have a derogatory effect on the final SST value that is reported. The **`Proximity_Confidence`** value is used in the assignment of sensor specific error statistics (SSES) that provide a bias and RMS error estimate for each pixel measurement.

The GDSv1 specifies separate schemes for infrared and microwave data streams that can be used to specify a proximity confidence value for each pixel. For SST data derived from infrared radiometers, sub-pixel cloud and wind speed are the main factors influencing the value that can be assigned to a **`Proximity_Confidence`** value. For SST data derived from microwave radiometers, proximity to rainfall flags and side lobe contamination in the proximity of land are the main factors influencing the **`Proximity_Confidence`** confidence value.

WP-ID2.1.1.14.1 Derivation of infrared proximity confidence value (IPCV)

The IPCV scheme proposed is based on that implemented at the EUMETSAT O&SI SAF (Brisson et al., 2001). The IPCV scale is shown in Table 2.1.1.14.1. IPCV values are assigned based on:

1. The spatial proximity of an SST observation to cloud contaminated areas (***D***). Clearly ***D*** will depend on the spatial resolution of the data stream in question and the geometry of the satellite sensor. For example, D values for geostationary observations of 4km will be different from polar orbiting radiometer data of 1.1km. 2 threshold values will be set by the GDSv1, ***D1*** and ***D2***, for each L2P data stream.
2. The temperature difference, ***ΔTmin*** defined in Eqn. 2.1.1.2. Four threshold values for ΔTmin will be set by the GDSv1 L2P configuration file.

Table 2.1.1.14.1 Infrared proximity confidence value (IPCV) scale definitions.

IPCV Value	Description
0	**Unprocessed:** Data that have not been classified
1	**Contaminated:** Data that that are most probably contaminated by cloud
2	**Bad:** Data that are probably contaminated by cloud
3	**Good:** Data that may be contaminated by cloud or influenced by a cool skin effect
4	**Acceptable:** Data that are reasonably distant from cloudy areas and in good agreement with the expected reference SST threshold
5	**Excellent:** Data that are far from any cloudy areas and in good agreement with the expected reference SST threshold
6	**Cool skin:** Data are far from any clouds but significantly cooler than the expected reference SST threshold. This could be isolated cloud, upwelling or other dynamical feature or a strong cool skin deviation.
7	**Failed:** Data are unusable due to one of the following reasons: 1. Cosmetically filled pixel 2. No L2 data 3. Value out of range 4. Sea Ice is present

Figure 2.1.1.14.1 presents a schematic diagram showing how IPCV values are assigned. In this figure, four ΔTmin thresholds (**`IPCVThresh1`** – **`IPCVThresh13`**) and two cloud proximity distances (**`IPCV_D1`** and **`IPCV_D2`**) delineate the set of data that correspond to the confidence scale provided in Table 2.1.4.1.1.

Threshold values for ΔTconfmin, (K)				
> **IPCVThres1**	3 (Acceptable)		5 (Excellent)	
< **IPCVThres1**		2 (Bad)		
> **IPCVThres2**			4 (Good)	
< **IPCVThres3**		1 (Cloudy)	6 (Cool skin or upwelling)	
	IPCV_D1	**IPCV_D2**	> **IPCV_D2**	D=Distance from cloudy pixel (km)

Figure 2.1.1.14.1 Schematic diagram showing how IPCV values are derived based on the deviation from a climatology of the coolest expected SST value and distance from cloud.

Note that the actual values for **IPCVThres1-3** and **IPCV_D1** and **IPCV_D2** are expected to vary for each sensor (Brisson et al., 2001) and the exact value of these parameters is a subject of on-going research and development within the GHRSST-PP. **IPCVThres1-3** and **IPCV_D1** and **IPCV_D2** values will be stored in the L2P configuration file that is specific to each RDAC or GDAC. The filename of this configuration file should be written into each L2P file using the optional global netCDF attribute "References" as described in Appendix A1.2.1. Finally, it is expected that the methodology used to derive IPCV may be modified in the GDSv2.

WP-ID2.1.1.14.2 Derivation of microwave proximity confidence value (MPCV)

The MPCV scheme proposed is based on the proximity of a pixel to the following criteria:

1. Physical distance from nearest rain flag (**MPCV_rain**)
2. Physical distance from nearest land (**MPCV_land**)
3. surface wind speed (**MPCV_wind**)

The MPCV scale is shown in Table 2.1.1.14.2.

Table 2.1.1.14.2 Microwave proximity confidence value (MPCV) scale definitions.

MPCV Value	Description
20	**Unprocessed:** Data that have not been classified
21	**Contaminated:** Data that that are most probably contaminated by land or rain
22	**Poor:** Data that may be contaminated by land or rain or may be influenced by wind speed extremes.
23	**Acceptable:** Data that are reasonably distant from rain and land flags and are within a favourable wind speed regime.
24	**Good:** Data that are good but may be influenced by seawater foam at higher wind speeds
25	**Excellent:** Data that are far from any land or rain and are within a favourable wind speed regime.
26	**Cool skin:** Data are far from any rain or land flags but are within a low wind speed regime where a significant cool skin layer may influence the observed SST.
27	**TMI:** data are from the TRMM TMI sensor. L2P confidence record flag TMI_flags

	should be consulted
28	**Failed:** Data are unusable due to one of the following reasons: 1. Cosmetically filled pixel 2. No L2 data 3. Value out of range 4. Sea Ice is present

Figure 2.1.1.14.2 presents a schematic diagram showing how MPCV values are assigned. In this figure, wind speed (**`MPCV_wind1`** and **`MPCV_wind2`**), the distance from land (**`MPCV_land1`** `and` **`MPCV_land2`**) and the distance from a rain flag (**`MPCV_rain1`** and **`MPCF_rain2`**) delineate the set of data that correspond to the confidence scale provided in Table 2.1.1.14.2.

Wind speed threshold value				
U < `MPCV_wind1`		**22 (Poor)**	**26 (Cool skin)**	
`MPCV_wind1` **< U <** `MPCV_wind2`		**23 (Acceptable)**	**25 (Excellent)**	
U > `MPCV_wind2`	**21 (Contaminated)**	**22 (Poor)**	**24 (Good)**	
	`MPCF_rain1`	`MPCF_rain2`	> `MPCF_rain2`	**R=distance from nearest rain flag**
	`MPCV_land1`	`MPCV_land2`	> `MPCV_land2`	**L=Distance from land**

Figure 2.1.1.14.2 Schematic diagram showing how MPCV values are derived based on the deviation from a climatology of the coolest expected SST value and distance from cloud.

Note that the actual threshold values shown in Figure 2.1.1.14.2 are expected to vary for a given sensor and the exact value of these parameters is a subject of on-going research and development within the GHRSST-PP. All threshold values will be stored in a L2P configuration file specific to each RDAC or GDAC that can be modified based on an assessment of sensor specific values. The filename of this configuration file should be written into each L2P file using the optional global netCDF attribute "References" as described in Appendix A1.2.1. Finally, it is expected that the methodology used to derive MPCV may be modified in the GDSv2.

WP-ID2.1.1.15 Assign L2P SSES error statistics

This test is used to set the L2P confidence data record variables **`bias`** and **`rms`** defined in Table A1.4.2 in Appendix 1. The GDSv1 specifies the following rules:

- **Rule 2.1.1.14a:** If native L2 pixel bias error and RMS error statistics are available, these values should be assigned to the L2P **`bias`** and **`rms`** variables respectively. In addition, the **`L2_native_bias`** and **`L2_native_rms`** flags of the corresponding L2P confidence data record should also be set.

If native L2 bias and RMS error values are not available they must be assigned based on the L2P **`Proximity_Confidence`** variable using sensor specific error statistics (SSES). On a regular basis (according to the volume of data accrued), SSES will be generated by an analysis of all available MMR data as described in WP-ID6. A mean bias and RMS error is derived for each sensor and a relationship between proximity confidence value and SSES is established. The L2P confidence data record variables **`bias`** and **`rms`** are then assigned the SSES bias and RMS value appropriate for the L2P **`Proximity_Confidence`** value. The GDSv1 specifies the following rule:

- **Rule 2.1.1.14b:** If native L2 pixel bias error and RMS error statistics are not available, these values should be assigned the most recent bias and RMS values of the SSES that is appropriate to the L2P confidence data record `Proximity_Confidence` value variables respectively. In addition, the `L2_native_bias` and `L2_native_rms` flags of the corresponding L2P confidence data record should be set to zero.

WP-ID2.1.1.16 SSI value assignment

This test is used to set the L2P confidence data record variable `SSI_value` defined in Table A1.4.2 in Appendix 1. An SSI measurement, coded according to the definition provided in Appendix 1 Table A1.4.2, nearest to the SST measurement in both space and time should be assigned to the L2P confidence data a record using the `SSI_value` field. These data are used by the GHRSST-PP Reanalysis Project and at later stages of the GDSv1. The GDSv1 specifies the following rules:

- **Rule 2.1.1.15a:** A L2P SSI measurement value, derived from satellite measurements, should be assigned to each pixel value using the `SSI_value` of the GDSv1 L2 confidence data record. The SSI measurement nearest in space and time to the L2 pixel SST value should be used.
- **Rule 2.1.1.15b:** If no SSI measurement is available, an SSI value should be used based on NWP output (See Appendix A3) nearest in space and time to the SST measurement should be used to set the `SSI_value` of the GDSv1 L2 confidence data record. A negative value should be set to indicate this value is not a measured value.

WP-ID2.1.2 Wind speed quality control tests

Wind speed measurements are required within the GDSv1 to derive L2P SST proximity confidence values (WP-ID2.1). The following QC checks should be applied to all L2 input wind speed data sets that are used at RDAC and GDAC to derive L2P wind speed data files. The tests applied are simple tests designed to verify that wind speed data are within realistic limits. RDAC and GDAC may choose to implement more sophisticated QC procedures if deemed necessary.

WP-ID2.1.2.1 Gross wind speed limit check

This test identifies wind speed values that fall outside of acceptable range of global wind speed values. The GDSv1 specifies the following rule:

- **Rule 2.1.2.1:** A valid wind speed observation must lie within a wind speed range of $0 < U <$ `HighWindSpeed`. Wind speed values that fall outside this range should not be used within the GDSv1.

The `HighWindSpeed` value is set in the L2P configuration file.

WP-ID2.1.3 Surface Solar Irradiance (SSI) quality control tests

SSI measurements are required within the GDSv1 to derive L3C SST diurnal confidence values (WP-ID3.2.4) and for assessing the magnitude and variability of significant diurnal SST variations. The following QC checks should be applied to all input L2 SSI data sets that are used at RDAC and GDAC to derive L2P SSI data files. These tests are simple tests designed to verify that SSI data are within realistic limits. RDAC and GDAC may choose to implement more sophisticated QC procedures if deemed necessary.

WP-ID2.1.3.1 Gross SSI limit check

This test identifies data that fall outside of acceptable range of global SSI values. The GDSv1 specifies the following rule:

- **Rule 2.1.3.1:** A valid SSI observation must lie within a SSI range of 0 < SSI < **`HighSSIValue`**. SSI values that fall outside this range should not be used within the GDSv1.

The **`HighSSIValue`** value is set in the L2P configuration file.

WP-ID2.1.4 In situ SST quality control tests

L2 in situ SST data are provided from a number of different sources and will be of variable quality. Some data providers (e.g., the French CORIOLIS and Canadian MEDS data centres) provide high quality QC'd L2 in situ data streams in near real time whereas L2 in situ data drawn down from the GTS must be QC'd locally before use. It is difficult to specify exact rules for the QC of L2 in situ data streams and each RDAC and GDAC centre is responsible for implementing basic L2 in situ QC procedures based on local knowledge and experience.

WP-ID2.2 Derivation of L2P confidence record `DV_confidence` and `DV_bias` values.

This test is used to set the L2P confidence data record diurnal variation confidence variable **`DV_confidence`** and **`DV_bias`** defined in Table A1.4.2 in Appendix 1. The **`DV_confidence`** value provides an assessment of the likelihood of any significant diurnal variability influencing the L2P data value. It is an important quality control parameter when considering bias correction of L2P data using in situ SST1m measurements and for the interpretation of L2P SST data values. The **`DV_bias`** provides an estimate of the magnitude of diurnal variation relative to the most recent SSTfnd data that are available and is used in the derivation of **`DV_confidence`**. It is defined according to Rule 2.2.1a:

- **Rule 2.2.1a:** the value of **`DV_bias`** will be calculated according to Eqn. 2.2.1:

DV_bias = SSTfnd – L2Pobs **(Eqn. 2.2.1)**

where SSTfnd[4] is the last available GHRSST-PP L4SSTfnd (explained in WP-ID4) value for a given L2P measurement defined in Appendix A3.1 and L2Pobs is the SST value of the L2P SST measurement.

The L4SSTfnd data product is produced by the GDSv1 system and represents the SST that is free of any diurnal effects (cool skin bias or warm stratified layer). **`DV_bias`** provides an indication of the sign and magnitude of diurnal variation, and thus also the location of any significant diurnal events. However, the interpretation of **`DV_bias`** in extremely dynamic areas of the ocean (e.g., W. boundary currents, coastal upwelling)

[4] SSTfnd in this case may be derived as a mean quantity over a window of several APPW.

requires careful thought as significant variability will be due to surface ocean dynamics in those areas.

The scheme that is proposed to derive **`DV_confidence`** values is shown in Figure 2.2.1 which are based the on various threshold values of surface wind speed, SSI and **`DV_bias`**. Wind speed and SSI are already part of the L2P confidence data and SSTfnd is provided as a reference data stream (see Appendix A3.1). Table 2.2.1 describes the scale of **`DV_confidence`** values used by the GDSv1.

	Time_1 02:00-sunrise	Time_2 Sunrise-12:00	Time_3 12:00-sunset	Time_4 Sunset-21:00	Time_5 21:00-02:00(T+1)
U_thresh_1 ($U<2ms^{-1}$)		SSI_1 SSI_2 SSI_3	SSI_1 SSI_2 SSI_3	SSIDay_1, uday-1 SSIDay_1,uday-2 SSIDay_2,uday-1 SSIDay_2,uday-2 SSI_3	SSIDay_1,uday-1 SSIDay_1,uday-2
U_Thresh_2 ($2 > U < 6\ ms^{-1}$)		SSI_1 SSI_2 SSI_3	SSI_1 SSI_2 SSI_3	SSI_1 SSI_2 SSI_3	
U_thresh_3 ($U > 6ms^{-1}$)					

Figure 2.2.1 Schematic diagram showing how L2P confidence record `DV_confidence` values are derived. All threshold values will be specified in a L2P_DV_confidence configuration file.

Table 2.2.1 L2P confidence record `DV_confidence` value definitions

Key for Figure 2.2.1	`DV_confidence` value	Description
	0	**Unprocessed:** Data are not classified due to lack of information (wind speed, SSI)
	1	**Persistent cool skin deviation** Wind speed is sufficient to inhibit the formation of thermal stratification or significant cool skin layers. Skin effect set to -0.15 K (Donlon et al. 2002)
	2	**Moderate cool skin deviation** A small cool skin deviation is possible
	3	**Weak diurnal warming** A warm layer is likely but is inhibited by wind mixing
	4	**Moderate diurnal warming** A warm layer is probable but may be limited by wind mixing
	5	**Strong diurnal warming** A significant warm layer is expected
	6	**Strong cool skin deviation** A significant cool skin deviation is expected
	7	**Insufficient information for derivation of DV_confidence**

The scheme shown in Figure 2.2.1 has been derived from extensive in situ measurements in the Arabian Sea, Western Tropical Pacific and NW Atlantic Ocean.

For the GDSv1, three wind speed thresholds (**`U_thresh1`**, **`U_thresh2`** and **`U_thresh3`**), three SSI thresholds (**`SSI_1`**, **`SSI_2`** and **`SSI_3`**), and five Time thresholds (**`Time_1`**, **`Time_2`**, **`Time_3`**, **`Time_4` and `Time_5`**) values are defined to delineate the set of data that correspond to the confidence scale provided in Table 2.2.1. **`uday-1`** and **`uday-2`** are mean wind speed thresholds calculated when SSI > 0. **`SSIDay-1`** and **`SSIDay-2`** are daily mean values for SSI when SSI > 0. Each threshold value is expected to vary for each sensor and the exact value of is a subject of on-going research and development within the GHRSST-PP. All threshold values will be stored in a L2P configuration file specific to each RDAC or GDAC. The filename of this configuration file should be written into each L2P data file using the optional global netCDF attribute "References" as described in Appendix A1.2.1. The following rule is specified for the computation of **`DV_confidence`**:

- **Rule 2.2.1b:** A value for the L2P diurnal **`DV_confidence`** variable defined in Table 2.2.1 will be assigned based on the scheme described in Figure 2.2.1.

Several other methods are available for the computation of **`DV_confidence`** ranging in complexity and input data requirements (e.g., Fairall et al. 1996; Kawai and Kawamura, 2002; Gentemann et al. 2002). All of these methods require validation and testing and it is expected that the methodology used to derive **`DV_confidence`** will be refined in the GDSv2 as more validation/diagnostic experience is gained within the GHRSST-PP.

WP-ID2.2.1 Estimation of L2P diurnal confidence `DV_shape` value

The L2P confidence data record variable **`DV_shape`** (defined in Table A1.5.2 in Appendix 1) provides a simple estimate of shape of diurnal variability at each grid cell based on an analysis of previous L2P SST data file grid cell confidence values.

Figure 2.2.2 shows a number of idealised diurnal variation (ΔTdv) "shapes" for a 24 hour period where ΔTdv has been calculated according to equation 2.2.1:

$$\Delta \mathrm{Tdv} = \mathrm{SSTsatellite} - \mathrm{SSTfnd} \qquad \textbf{(Eqn 2.2.1)}$$

where **SSTsatellite** is the measured SST from either a microwave or an infrared sensor. The aim is to establish an estimate of both the diurnal bias and the shape of the diurnal SST signal referenced to SSTfnd.

In Figure 2.2.2, Shape 1 (blue) depicts a situation with no diurnal stratification typically found in moderate to high winds and/or very cloudy conditions (U > 8 m/s). Shape 2 (black) is an ideal shape for a steady 2 m/s < U < 8 m/s and with no dramatic variations in wind speed. The maximum magnitude and gradient of the diurnal shape will vary according to the intensity of the wind speed and SSI. Shape 3 (green line) describes a situation with low wind in the morning (U < 2 during 08:00-12:00h) characterised by a sharp initial heating gradient. In this scenario, the wind speed then increases (2 m/s < U < 6 m/s) in the afternoon (indicated by a green arrow) and a weakening the heating gradient. Significant stratification is expected in the upper 1m of the water column. Shape 4 (yellow line) describes the scenario for a moderate wind in the morning (2 < U < 6 m/s) that diminishes to a low wind speed in afternoon (U < 2 during 12:00h-16:00 h) with a moderate wind at night. The increase in wind speed is indicated by a yellow arrow. Shape 5 (red line) considers the scenario with low wind speeds (U < 2m/s) throughout the day. This scenario leads to the strongest stratification in upper 1m that persists through out the night. Shape 6 (light blue line) describes the scenario in which a wind burst (increase) occurs in the afternoon (indicated by a light blue arrow) which breaks down any stratification that had occurred in the morning. Finally, Scenario 7 (pink line) considers a wind burst in the morning (indicated by a pink arrow) with a decrease in wind speed in the afternoon. In this situation there is still time for relatively strong stratification to develop especially if U remains low.

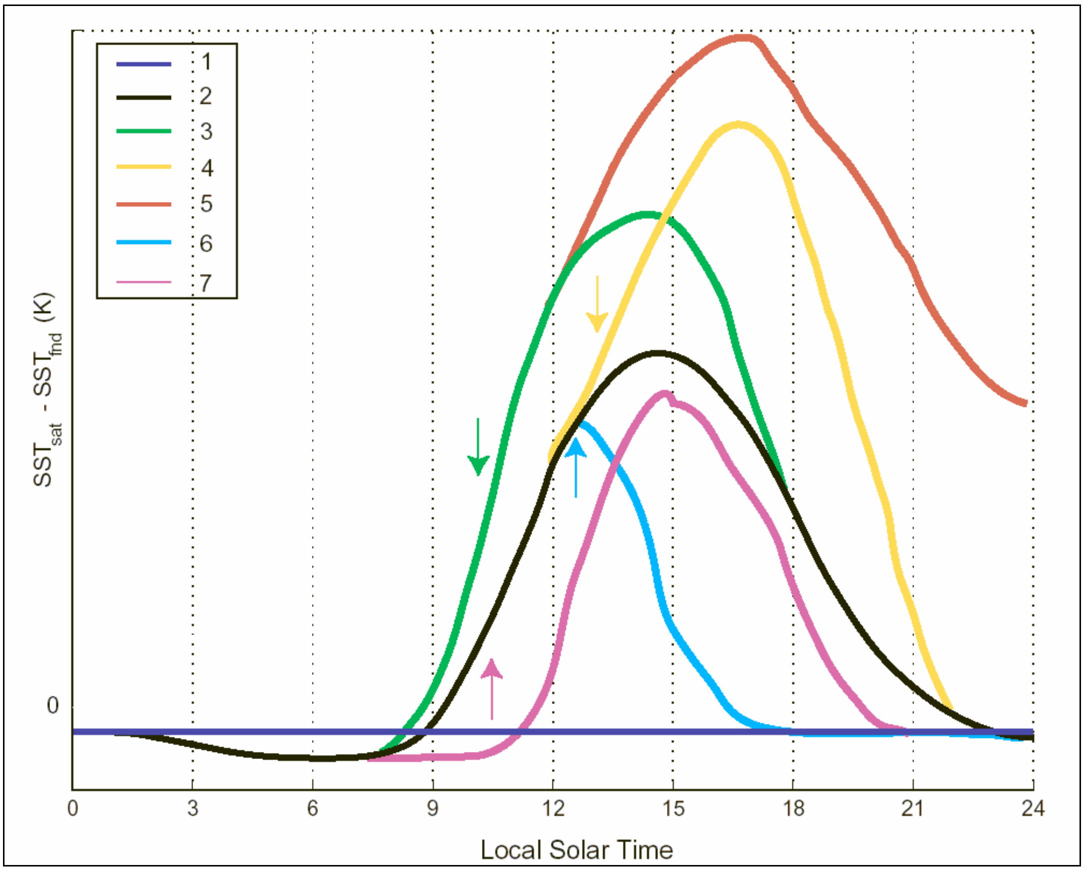


Figure 2.2.2 Idealised diurnal variation ($SST_{0.2m}$-SSTfnd) signals for a 24 hour period. The wind event refers only to the example type 3 curve. See text for a full description of this figure.

In nature there is considerable variation in the magnitude and phase of diurnal SST signals according to amongst many things, the variability of incoming SSI, surface wind speed (mixing), current structure, and water quality. As such, figure 2.2.2 provides a framework that can be built upon to provide phase and magnitude estimates of diurnal variability.

Figure 2.2.3 provides a set of example observations that underpin the generalisations made in Figure 2.2.2 taken in the NE Atlantic, Arabian Sea and Tropical Western Pacific. Several stratification scenarios are shown according to specific wind speed regime (also shown). While these examples serve to justify the choice of diurnal variation shapes shown in Figure 2.2.2, it should be immediately clear that complex variations are to be expected in nature. For example, in the top right hand panel (low wind speed at night), warm stratified layers are not broken down and a significant deviation persists between the SST above a depth of 5m. In the example provided for low wind speed (centre top), even small wind bursts are seen to have an effect on the shape of diurnal SST variation, albeit quite small.

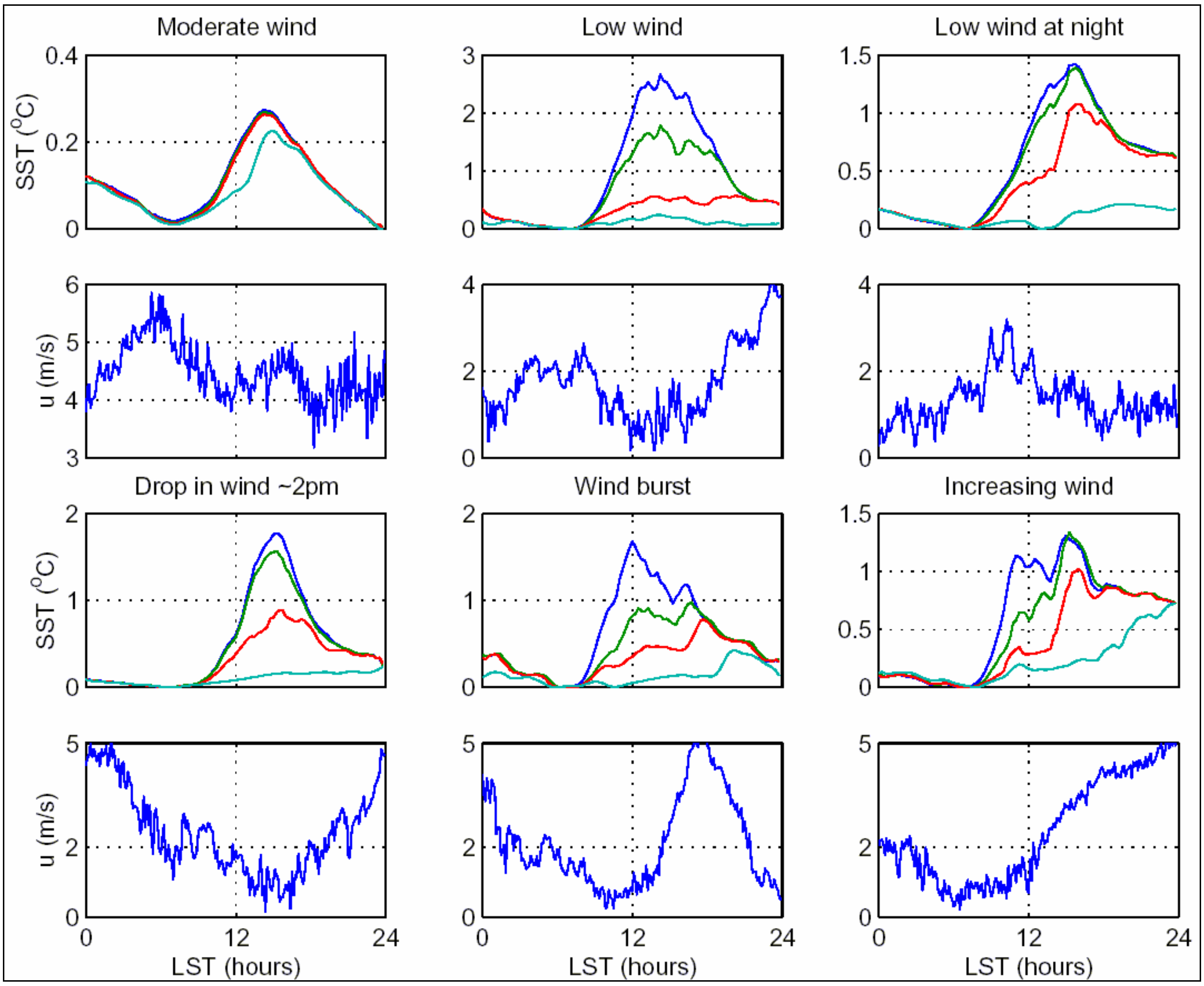


Figure 2.2.3. Diurnal signals (referenced to 10m depth) for various wind speed regimes plotted based on in situ buoy observations in the Arabian Sea, Western Pacific and NW Atlantic Ocean. Colours indicate the depth of SST measurement: Blue=0.2m, Green=1.0m, Red=2.5m And cyan=5.0m. (A. Stuart Menteth)

RDAC and GDAC are encouraged to assign a value for **`DV_shape`** based on the code values provided in Table 2.2.2 and using the available **`DV_confidence`** data to estimate **`DV_shape`**. RDAC and GDAC are also encouraged to use parameterisations of diurnal variation derived from NWP and satellite observations (e.g., Kawai and Kawamura, 2002; Fairall et al. 1996; Gentemann et al. 2002) to establish values for **`DV_shape`**, **`ΔTdv`** and **`DV_confidence`**.

Without further research, development and validation, it is not possible to specify exactly which method will provide the best estimate of diurnal SST variation. Consequently, it is foreseen that the methods used to characterise diurnal SST variability will be significantly upgraded in the GDSv2 based on further research and development within the GHRSST-PP.

Table 2.2.2 L2P `DV_shape` value scale.

`DV_shape` value	Description
0	`DV_shape` not evaluated
1	Steady high wind speed (u) situations where u > 10 ms^{-1}. At night a constant cool skin deviation of -0.17 K is maintained and during the day no significant warm layer is expected (Donlon et al ., 2001). No significant variation in Δ**Tdv** is expected.
2	Steady light wind conditions (u < 2ms^{-1}) and clear skies prevail allowing strong SSI flux or long wave cooling flux at night. A strong diurnal signal (cool skin or warm layer) is highly likely to form in these conditions.
3	Low wind in the morning (U < 2) (00:08h-12:00h) results in a sharp heating gradient. In the afternoon wind speed increases (2 < U < 6 m/s) resulting in a weaker heating gradient. Stratification is expected in the upper 1m during the day.
4	A moderate wind speed prevails in the morning (2 < U < 6 m/s) diminishing in the afternoon (U < 2 m/s during 12:00h-16:00h). A moderate wind speed prevails at night. Stratification is expected in the day with a moderate cool skin at night.
5	Low winds prevail all day (U < 2m/s) and strong stratification is expected in the upper 1m. Stratification may persist well into the night.
6	A wind burst (increase) occurs in the afternoon that breaks down stratified layers.
7	A wind burst (increase) in the morning followed by a drop in wind speed in the afternoon results in a stratified layer developing (especially if U < 2 m/s).

WP-ID2.3 Evaluate L2P data

Each RDAC and GDAC center will evaluate whether or not a L2P data set should be admitted to the GDS processor based on objective criteria.

WP-ID2.3.1 Reject L2P data file

If a L2 data file has been rejected by an RDAC/GDAC, an error must be raised and logged at the ERRLOG system as described in Appendix A7.2. In addition, an e-mail message should be sent to the L2 data provider that explains that the L2 data has been rejected form the GHRSST-PP GDS stating the reason why. A copy of this e-mail should be sent to the ERRLOG system log. In addition, an appropriate OPLOG L2_reject_DPSR should be submitted to the OPLOG log as described in Appendix A7. Note that no MMR-FR metadata are prepared for rejected L2P data streams.

WP-ID2.3.2 Accept L2P data file

A L2P data file and corresponding MMR-FR metadata record should be generated as specified in WP-ID2.3 and WP-ID2.4 respectively.

WP-ID2.4 Format L2P data file

Each L2P data set will be formatted and archived as a netCDF file following the data format described in Appendix A1.2 and A1.4.

WP-ID2.5 Generate and register L2P MMR-FR metadata record

For each L2P data a corresponding MMR-FR must be prepared by the responsible RDAC/GDAC processor and registered at the MMR.

WP-ID2.5.1 Format L2P MMR-FR metadata record

A MMR-FR metadata record should be prepared for each L2P data file according to the specification provided in Appendix A6.2.

WP-ID2.5.2 Register L2P MMR-FR metadata record with MMR system

MMR-FR metadata records should be delivered to the MMR system as soon as possible after the L2P data set has been evaluated following the procedure described in Appendix A6.4.

WP-ID2.6 Submit an entry to the OPLOG log system

In order to maintain a record of L2P data processing, an L2P_DPSR entry within the OPLOG log should be made via e-mail as explained in Appendix 7.

WP-ID3 Derivation of Single Sensor Error Statistics (SSES)

WP-ID3	**Derivation of Single Sensor Error Statistics (SSES)**
Work Package number :	**WP-ID3**
Leader:	**TBD**
Aim:	**To specify the procedures and methods that will be used to derive single sensor error statistics (SSES) for each L2 satellite SST data stream**

1	**Objectives:** 1. To specify the structure of the GHRSST-PP Match up database (MDB) of near contemporaneous satellite and in situ data. 2. To specify the format of GHRSST-PP validation records that collectively form the GHRSST-PP match up data base (MDB). 3. To specify the procedure for generating a MDB data records at GDAC and RDAC 4. To specify the derivation of SSES using MDB data records
2	**Description:** This WP describes the GHRSST-PP MDB and the data records that should be produced at each RDAC/GDAC containing in situ SST and other ocean-atmosphere data as available that are matched to near contemporaneous L2 SST measurements. It is recognised that in situ observations are (a) scarce and (b) will not be collected by dedicated GHRSST-PP measurement campaigns. The GHRSST-PP will use data from the GTS and from regional data providers in near real time that provide QC in situ SST data (e.g., MEDS, CORIOLIS, JMA). The GHRSST-PP MDB provides a data resource that can be used to derive an estimate of the mean bias and RMS error associated with each L2 data stream. The GDSv1 requires that error estimates are assigned to every L2 satellite SST measurement during the derivation of L2P data products. As an in situ observation is not available for every L2 SST measurement, single sensor error statistics (SSES) are derived by a statistical analysis of all MDB data records for a given sensor type and period. SSES consist of a statistical mean bias and RMS error estimate that are correlated to each mean L2P **`Proximity_confidence`** value assigned to every L2P measurement in WP-ID2. In this way, the limited amount of in situ data may be utilized to maximum benefit. A precedent for this approach has been set by the EUMETSAT O&SI SAF and the US Navy NAVOCEANO operational SST data production systems as discussed at the third GHRSST-PP workshop.
3	**Inputs:** 1. L2P data files 2. In situ ocean-atmosphere measurements from buoys, drifters and ships 3. SSES configuration file
4	**Outputs:** 1. MDB records delivered to the GHRSST-PP MDB 2. SSES for each sensor derived on a on a regular basis used in WP-ID2
5	**Acceptance tests:** 1. MDB records are formatted correctly and are delivered to the GHRSST-PP MDB system 2. SSES analysis is performed on a weekly basis and SSES are published to all RDAC and via the GHRSST-PP UIS

6	**Metric for performance assessment:** 1. MDB records are produced in a timely manner 2. MDB records are provided and ingested at the MMR in near real time. MDB rejection rate < 2% 3. OPLOG log files are correct and timely 4. SSES are available in a timely manner (several days) at regular intervals

This work package is dedicated to the generation of satellite SST matched to in situ observations that are used to provide a quantitative error estimate for each L2P satellite SST measurement. The match up data are stored in a GHRSST-PP Match Up Database (MDB) that forms a core data resource for both the derivation of single sensor error statistics (SSES) and for the on-going independent validation of satellite data streams described in WP-ID5. The MDB is a collective resource that is shared and populated by all RDAC and GDAC within the GHRSST-PP and allows an independent assessment of the absolute error of all L2 satellite data streams during the derivation of L2P data products in WP-ID2. This WP describes the derivation of MDB data records and SESS whereas WP-ID5 is concerned with the more general issue of GHRSST-PP data product validation. Both WP rely on the regular population of the GHRSST-PP MDB of using near real time satellite and in situ observations.

While the core of the MDB is focussed on satellite-to-in situ match up data records, in some cases it may be appropriate to match satellite data to satellite data (MDB-FR formatted according to Appendix A4.2). Thus, two types of MDB record will be produced by this work package: satellite-to-in situ match-ups and satellite-to-satellite match ups, the latter forming an important diagnostic for the assessment of the relative bias differences between L2 input SST data streams.

Figure 3.1.1 presents a functional breakdown diagram of the processing tasks within WP-ID3 and clearly identifies two interrelated components:

1. The derivation of MDB records and their storage in a database system
2. The derivation of SSES using MDB data records

GHRSST-PP L2P data files and in situ observations form the input into this work package that are initially matched in space and in time (WP-ID3.1.1) for each individual satellite sensor. If in situ data can be matched with satellite data then a MDB data record is prepared and formatted according to the specification set out in Appendix A4.1 (WP-ID3.1.2). Each MDB data record is stored as a file record (MDB-FR) that is associated with a data set description record (MDB-DSD) in a similar manner to the GHRSST-PP Metadata repository system. For a given satellite sensor there will be a unique MDB-DSD containing a description of the L2 satellite data stream and many associated MDB-FR containing the MDB data records themselves.

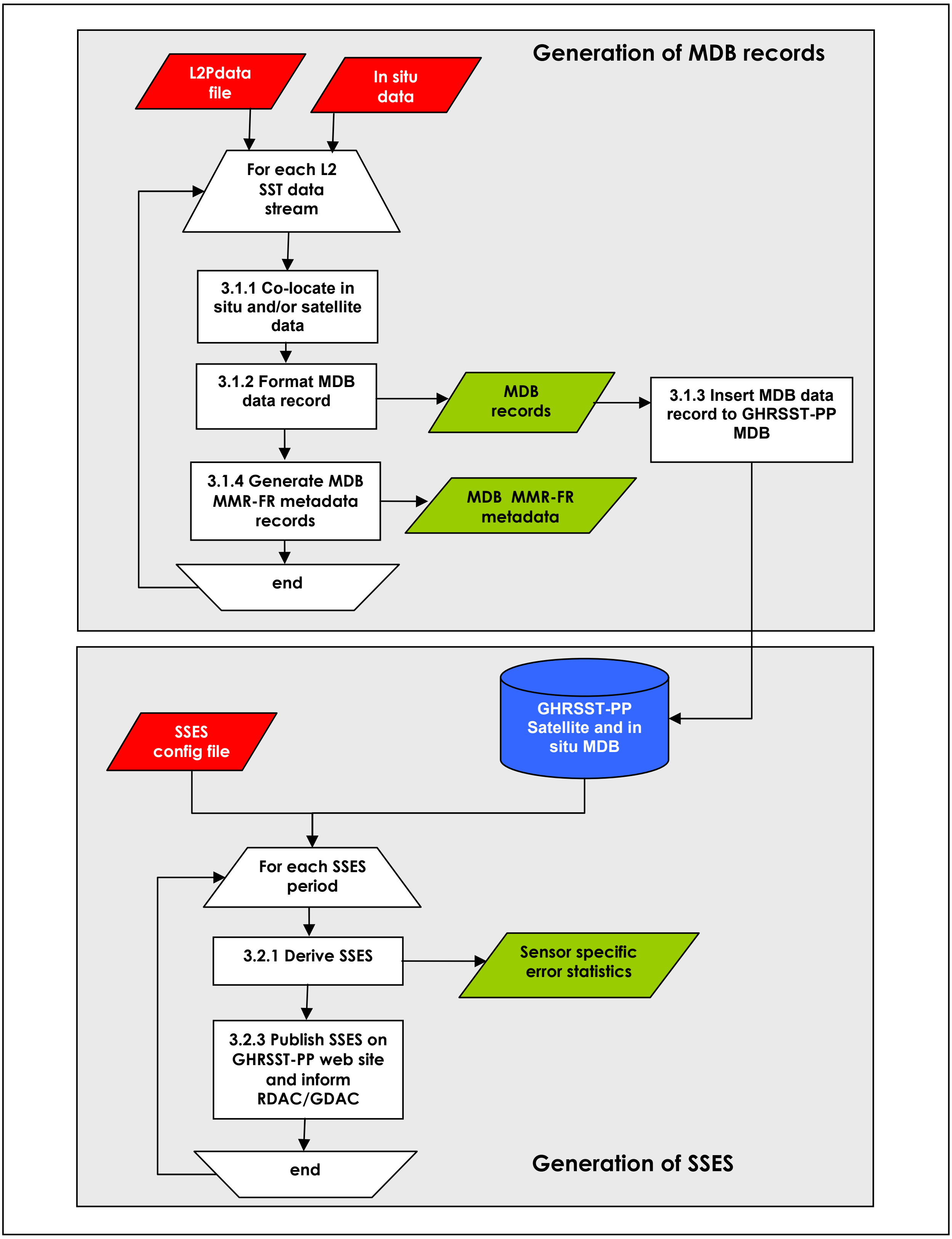


Figure 3.1.1 Functional breakdown diagram of WP-ID3.

Each MDB-FR record is then ingested into the GHRSST-PP MDB (WP-ID3.1.3) and a corresponding MMR metadata record is prepared and submitted to the GHRSST-PP MMR system according to the format provided in Appendix A6. MMR data records are used to search[5] for specific MDB-FR entries using the same interface that is used to discover all GHRSST-PP data products.

At regular intervals, known as SSES periods, an analysis procedure is used to derive mean bias and RMS values for each mean L2P data stream **`Proximity_confidence`** value (WP-ID3.2.1). SSES provide independent error quantification for each L2 data stream. SSES are published on the GHRSST-PP web site as part of the GHRSST-PP User Information Service (UIS) and all RDAC/GDAC are informed of the latest issue of SSES fir use in the generation of GHRSST-PP L2P data products. A SSES configuration file is used to control the derivation of SSES statistics for each L2P data stream.

WP-ID3.1 The GHRSST-PP Match Up Database (MDB) system

The GHRSST-PP MDB system is foreseen as a relational database containing all GHRSST-PP MDB data records. A precedent for this system has been set by the Miami Pathfinder Match –up database system that is fully described in Podesta and Evans (2003). The GHRSST-PP MDB will be maintained at the GDAC using a similar database system to the GHRSST-PP metadata repository (MMR) although RDAC are encouraged to maintain their own MDB systems with the assumption that at regular intervals (before the derivation of SSES) RDAC data are copied into the central GHRSST-PP MDB system.

Access to the MDB file record data (MDB-FR) will be unrestricted to all participants of the GHRSST-PP but will be subject to the data access/usage requirements and stipulations of data providers (in some situations, match up data records are sensitive data e.g., shortly after the launch and commissioning of a satellite sensor). General access to the MDB will require that users are registered and agree to any data access and usage agreements that are relevant to MDB records.

MDB-FR data records should be produced and ingested at the MDB in near real time (1-7 days) in order for SSES to be generated in a timely manner. Timeliness of MDB data is related to the availability of in situ data that itself, requires careful quality control and validation. Collaboration with operational data centres that are well practiced in the QC of in situ observations should be established and maintained by RDAC and GDAC especially in the case of data that are made available on the GTS. The GHRSST-PP expects to work closely with the CORIOLIS data centre (France), the Japanese Meteorological Agency (JMA), and the MEDS data centre (Canada) as described in Appendix 3.4.5 and Appendix 3.4.6.

3.2.1 WP-ID3.1.1 Derivation of MDB data records

The primary data inputs to GHRSST-PP MDB records are L2P data streams and independent in situ observations. For an MDB data record to be accurate and legitimate the comparison must be between like measurements. Since satellite-to-in situ and satellite-to-satellite data are spatially and temporally inhomogeneous, this implies that limits of coincidence must be applied to limit the contributions to the error budget of the validation exercise from these sources to acceptable levels (say <0.1K)

In the case of in situ-to-satellite data match ups, the following GDS-v1 rules are specified:

---

[5] The MDB will be a relational database system and may be searched directly using SQL syntax as required. However, experience has shown that a WWW type interface with advanced by structured "standard" query commands is preferred by many users.

- **Rule 3.1.2a:** The spatial separation of the validating measurement and the centre point of a satellite-based measurement (centre of pixel) must not be more than 3 km (2 times the error of a well mapped high resolution (1km) satellite image) in the case of high resolution (~1km at nadir) L2 data streams. In the case of low resolution (e.g., 25 km gridded microwave SST measurements) the threshold may be relaxed to 10 km. In the latter case, a transect of several measurements obtained within the satellite FoV will provide a better representation of the in situ SST and are preferred to single point observations from moored or drifting buoys.
- **Rule 3.1.2b:** Adjacent pixels or cells (5x5 array) of the satellite data, must be extracted to provide information on the spatial variability in the vicinity of the validation point. Appropriate 'missing data' values will be used if any of the elements of the pixel array extend beyond the edges of the instrument swath.
- **Rule 3.1.2c:** All L2P confidence data each pixel stored in a MDB-FR must accompany the extracted data.
- **Rule 3.1.2d:** Because of the day-night asymmetry in the growth and decay of the diurnal thermocline, more stringent conditions for acceptable time intervals are required during the day than at night. Daytime validation data should be within 30 minutes of the satellite data, and within 2 hours at night. For ambiguous cases, such as close to the terminator, the shorter time interval should be used.
- **Rule 3.1.2e**: MDB data obtained in regions characterised by dynamic SST fields should be labelled as such in the associated MMR-FR metadata record. Validation data obtained in these regions should be treated with caution.

The timeliness of the validating data is important, but speed of delivery should not be used as an excuse to compromise the quality of the measurements. Where possible quality assured validation data should be delivered to the RDAC within the current Processing Window, but this may not be achievable in most cases. Certainly a delay of 24-72 hours is to be expected when using data from operational in situ data centres due to the latency incurred by QC procedures.

Additional geophysical fields will be required to provide insight into any uncorrected perturbations to the SSTs which might result in systematic errors. These should include surface wind, atmospheric pricipitable water and an atmospheric aerosol index. Ideally these should be derived from contemporaneous measurements of other sensors on the same spacecraft as the SST sensors, but if this is not possible, data from the same day and from different satellite sensors (and therefore at different times) are acceptable. However, given the scarcity of in situ observations, it is acceptable to create MDB records without such additional data.

In the case of satellite-to-satellite match up data, different rules must be specified as some sensors have a finer spatial resolution than others, as shown in Figure 3.1.1.1 For these cases the rule will be to average the values of all pixels which overlap the lower resolution pixel entirely and which have a L2P confidence record **`Proximity_Confidence`** value equal to the highest encountered within the cell, to produce a single value.

The following rule is specified for this purpose:

- **Rule 3.1.3:** When matching a fine resolution data stream to a corresponding coarse resolution data stream (e.g. AATSR data matched to AMSRE), corresponding match up values are derived as an average of the fine resolution data that completely overlap the coarse resolution pixel size. Only fine resolution data having a L2P `Proximity_Confidence` value equal to the highest encountered within the coarse resolution FoV should be used to produce a single SST value.

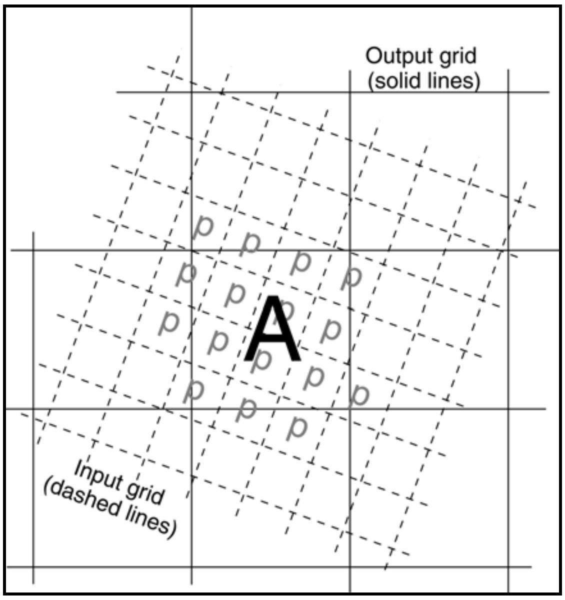


Figure 3.1.1.1 To illustrate the approach when high resolution satellite SST data are to be matched to low resolution satellite SST data All pixels labelled p in the high resolution data are possible contributors to match up to the coarse resolution satellite SST data (labelled A).

WP-ID3.1.2 Format of MDB File Records (MDB-FR)

Each satellite SST matched to in situ SST MDB-FR will be formatted and archived as an ASCII text file following the data format described in Appendix A4.1 Appendix A4.2 describes the MDB-FR format for satellite SST matched to satellite SST.

WP-ID3.1.3 Insert MDB record into the GHRSST-PP MDB system.

MDB-FR data records should be sent to the GHRSST-PP MDB as an e-mail message in ASCII text format with 1 field per line according to the specification provided in Appendix A4.1 and A4.2 following the XML DTD provided in Appendix A4.3. These messages will be parsed and inserted into the GHRSST-PP MDB if correctly formatted. An error message will be returned to the data provider if the MDB-FR is incorrectly formatted.

WP-ID3.1.4 Generate and register MDB MMR-FR metadata record

For each MDB data record, a corresponding Master Metadata Repository File Record (MMR-FR) must also be prepared by the responsible RDAC/GDAC processor and registered at the MMR. These data are used to search the MDB system for appropriate MDB-FR for a given time, location and sensor name (for example). MDB-FR will be registered at the MMR under a unique MDB MMR Data Set Description record (MMR-DSD).

WP-ID3.1.4.1 Format MDB MMR-FR metadata record

A MMR-FR metadata record should be prepared for each MDB-FR data record according to the specification provided in Appendix A6.2.

WP-ID3.1.4.2 Register MDB MMR-FR metadata record with MMR system

MMR-FR metadata records should be delivered to the MMR system as soon as possible after the MDB-FR data record has been derived following the procedure described in Appendix A6.4.

WP-ID3.2 Derivation of sensor specific error statistics (SSES) using MDB

At regular SSES periods (weekly, monthly or quarterly, TBC) SSES will be derived for each **`Proximity_confidence`** value (1-7) and for each of the L2 data streams indicated in Table A3.2. SESS form an important input to WP-ID2 where a SESS value is assigned to each L2 satellite SST measurement based on the **`Proximity_confidence`** value that has been assigned to each measurement.

SSES are derived in the following way:

1. All MDB-FR for a given sensor and SSES period are extracted from the GHRSST-PP MDB system.
2. The extracted data are separated according to the **`Proximity_confidence`** value that has been assigned to the L2P data.
3. The data are then averaged and the mean bias and RMS difference (Satellite minus in situ) is computed for each **`Proximity_confidence`** value.
4. In the case where insufficient data are available within a given SSES period to calculate a significant mean and RMS, data from previous SSES periods may be used.

It is expected that the derivation of SSES will evolve as the GHRSST-PP science Team becomes more experienced with the derivation and application.

WP-ID4: Generation of L4 Analysed data products (L4*)

WP-ID4 Generation of Analysed data products	
Work Package number :	WP-ID4
Leader:	Hiroshi Kawamura (kamu@ocean.caos.tohoku.ac.jp) Pierre LeBorgne (LeBorgne@meteo.fr) and Richard Reynolds (Richard.W.Reynolds@noaa.gov).
Aim:	To specify the data format, processing method and algorithm that will be used to generate L4 analysed data products (L4*)

1

Objectives:

1. Specification of the data processing method and algorithm that will be used to generate L4 SST1m (L4SST1m) data products
2. Specification of the data processing method and algorithm that will be used to generate L4 SSTfnd (L4SSTfnd) data products
3. Specification of the L4* MMR-FR metadata that should be prepared and registered at the MMR
4. Specification of HR-DDS data granules and associated HR-DDS MMR metadata that should be extracted form each L4* data product

2

Description:

The GDSv1 will produce several different SST data products using analysis procedures that take L2P data products as input. GDS L4*[6] data products provide complete coverage SST data fields that are free of gaps for each APPW. Two L4 data products will be provided by the GDSv1[7]:

1. SST1m using the New generation SST (NGSST) methodology
2. SSTfnd including a diurnal variation providing information describing the location, phase and magnitude of SST diurnal variations

This work package specifies the "rules" and procedures that will be used to generate L4 data products, register each product at the MMR and, extract HR-DDS data granules and associated metadata.

3

Inputs:

1. L2P data files
2. L4 data processor configuration file

4

Outputs:

1. L4SST1m data product (L4SST1m) and associated confidence and error statistics for each APPW
2. L4SSTfnd data product (L4FND) and associated confidence and error statistics for each APPW
3. MMR-FR metadata record for each L4* product for each APPW
4. Extracted HR-DDS data granules and HR-DDS MMR-FR metadata record for each data granule

5

[6] L4* refers to all L4 data products throughout this document.
[7] It is foreseen that new and revised L4 data products may be provided by the GDSv2 based on the experience of the GDSv1.

	Acceptance tests: 1. Identical L4* output data products are produced at each RDAC and GDAC when the GDSv1 test data set us used compared to the reference processor. 2. L4* MMR-FR metadata entries are timely and accurate. 3. HR-DDS L4* data products are extracted correctly 4. L4* HR-DDS-MMR –FR metadata is correct and accepted by the HR-DDS MMR system
6	**Metric for performance assessment:** 1. L4* products and HR-DDS L4* granules are produced in a timely manner 2. L4* MMR-FR metadata records provided and ingested at MMR and HR-DDSS MMR in real time. MMR and HR-DDS MMR rejection rate < 2% 3. OPLOG L4_DPSR records correct and timely

This part of the GDS-v1 describes the rules and methods used to generate L4* analysed data products. Two L4 data products are specified by the GDSv1: an estimate of SST1m based on the New generation SST (NGSST) methodology (Guan and Kawamura, 2002), an estimate of the SSTfnd that accounts for both warm layer and cool skin effects including an estimate of the SSTskin at 6 hourly intervals that can be used together with the SSTfnd or SST1m product to estimate the diurnal variation throughout an APPW. All L4 data products will be produced for each APPW and made available to the user community in real time. These data products are based on user requests that are described in the GDIP and to provide some flexibility for developing the L4 data processor.

Figure 4.1 presents a functional breakdown diagram of the data processing steps required to produce L4* data products for each APPW. The processor takes L2P data files as input together with a L4 configuration file containing threshold and data processor configuration settings. For each APPW the L4SST1m and L4SSTfnd are calculated using an optimal interpolation (OI) procedure for each output grid cell (WP-ID4.1, WP-ID4.2 and WP-ID4.3 respectively in Figure 4.1). Each L4* data product is then formatted as a L4* data file (WP-ID4.4 in Figure 4.1) and a separate MMR-FR is prepared and submitted to the MMR (WP-ID4.5 in Figure 4.1). HR-DDS granules are extracted from each L4* data product (WP-ID4.6 in Figure 4.1) and a HR-DDS MMR-FR prepared and submitted to the HR-DDS MMR system (WP-ID4.7). Finally a L4_DPSR record is prepared and submitted to the OPLOG data log system (WP-ID4.8).

It is expected that L4* data products will evolve with experience based on the results obtained from validation studies and analysis of independent SSES and HR-DDS granules together with in situ observations.

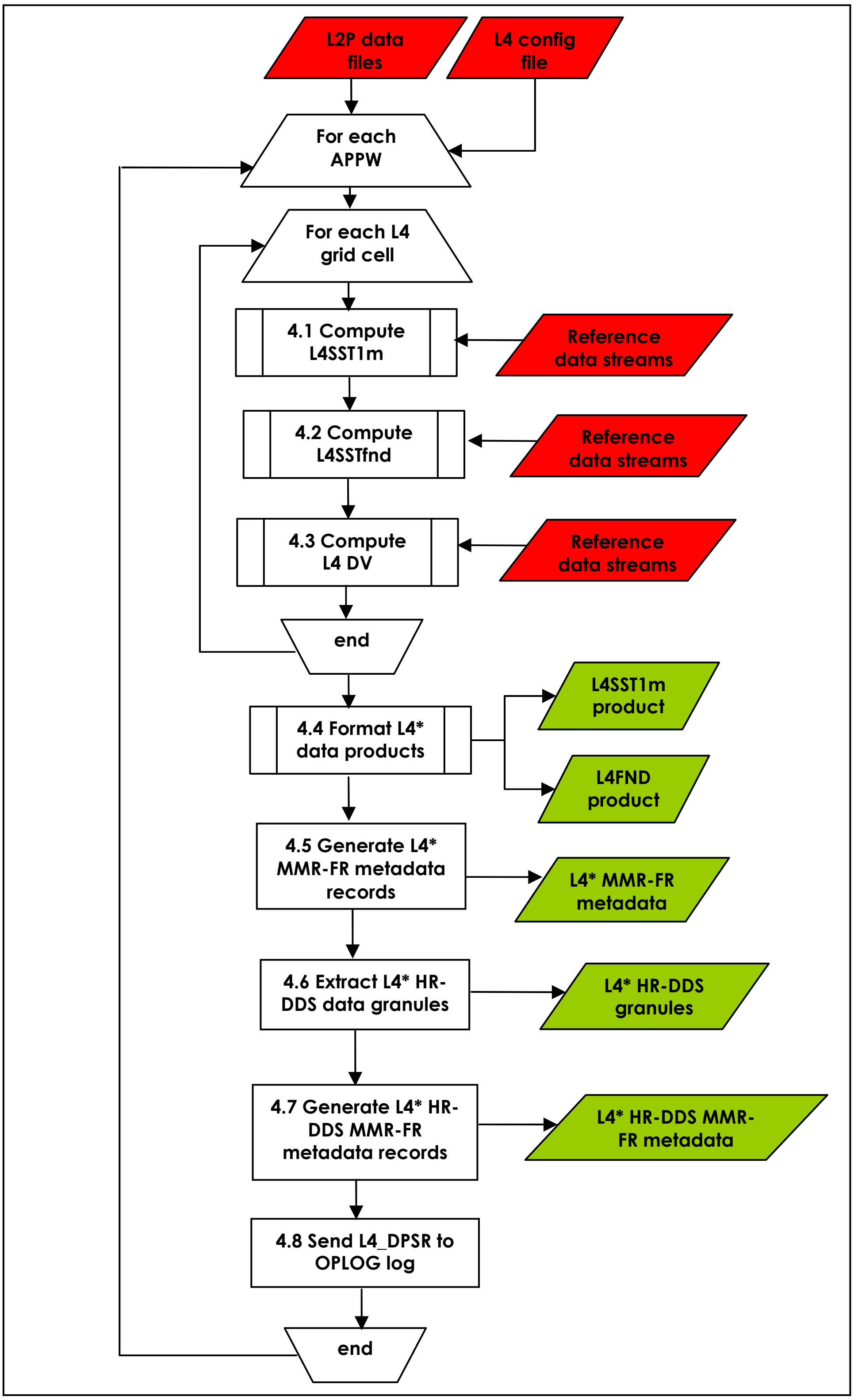


Figure 4.1 Functional breakdown diagram for WP-ID4 showing the data processing steps required to generate GDSv1 L4* SST data products.

WP-ID4.1 Compute L4 SST1m data products (L4SST1m)

An estimate of the SST1m will be generated by the ISDI-PMv1 to produce L4NGSST data products following the New Generation SST (NGSST) scheme developed at the University of Tohoku, Center for Atmospheric and Oceanic Studies (Guan and

Kawamura, 2002). The NGSST methodology accounts for thermal stratification effects in conditions of calm-winds and high SSI. The NGSST provides an estimate of the SST1m based on a combination of SST from infrared sensors and microwave sensors. Satellite SSTs are merged by using a correlation-based objective analysis to generate a cloud-free, quality-controlled, and daily SST product (with spatial resolution of 0.05-degree tested at the RDAC). Recently, regional/seasonal dependent decorrelation length scales have been developed for the objective analysis procedure.

Figure 4.1.1 presents a schematic overview of the data flow within the NGSSTv2 processing scheme.

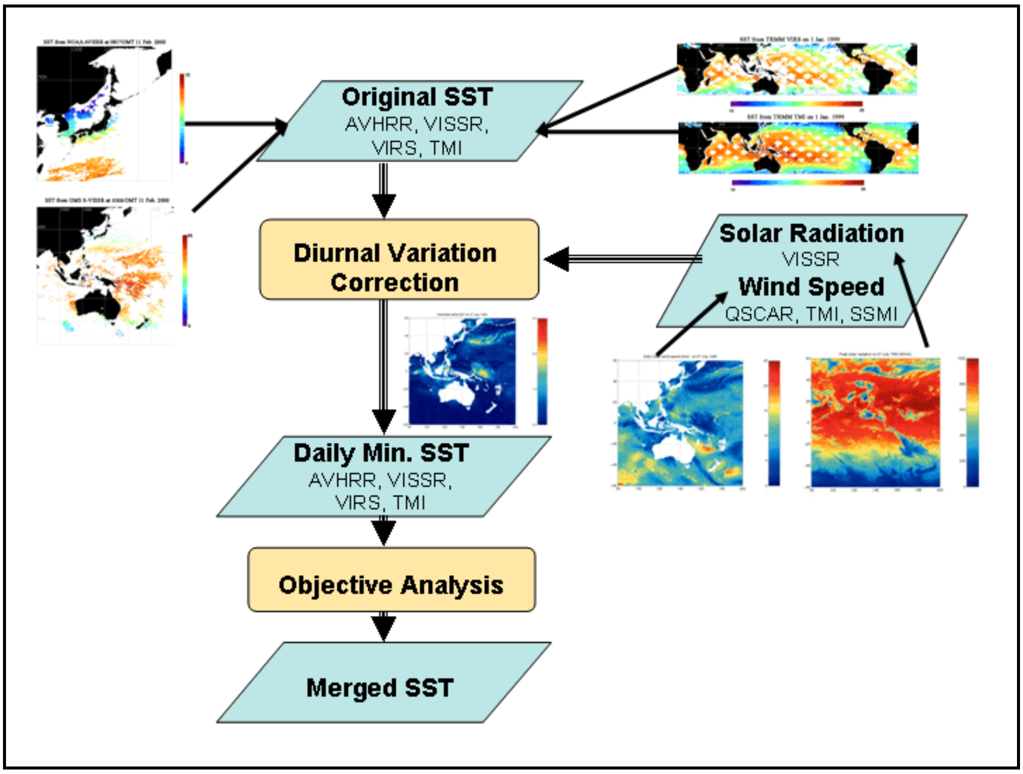


Figure 4.1.1 A schematic overview of the New Generation SST version 2 algorithm

In the NGSST processing scheme, a diurnal variation bias correction is first applied to infrared and microwave SST retrievals. Each SST is normalised to a daily minimum SST using a diurnal variation amplitude bias value estimated using a semi empirical model that takes as input satellite SSI and wind-speed (Kawai and Kawamura, 2002). Finally, an objective analysis method is applied to derive a complete SST field (Guan and Kawamura, 2002). Although the NGSST algorithm provides an estimate of the SST1m, at this depth significant diurnal variation of temperature under conditions of calm winds and high SSI is expected.

This section of the GDS summarises the NGSST method and describes the implementation of NGSST v1 and v2 at the Japanese RDAC.

WP-ID4.1.1 NGSSTv2 bias correction strategy (diurnal variation correction)

Tanahashi et al. (2002) demonstrated that the SST diurnal variations (DV) of SST within the footprint of the Japanese Geostationary Meteorological Satellite (GMS) can be estimated using SSI computed from GMS and satellite-derived wind-speeds together with a semi-empirical model DV (Kawai and Kawamura, 2002).

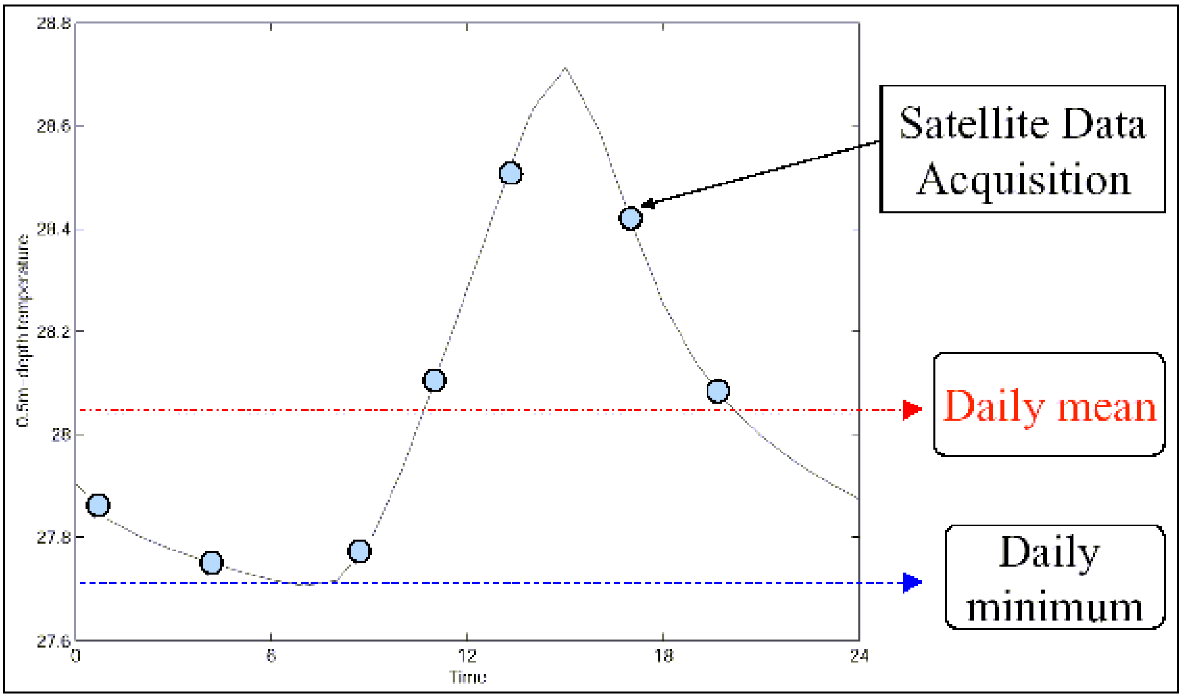


Figure 4.1.1.1 NGSST diurnal variation definitions. The black dotted line indicates an idealised diurnal variation of SST for a 24 hour period starting at local midnight. The SST is sampled by several satellite sensors (blue circles). The daily mean SST and daily minimum SST values are marked.

Figure 4.1.1.1 shows a schematic diagram identifying the major elements of the NGSST bias correction methodology. In this figure the SST for a given point has been plotted as a time series over a 24 hour period and shows an idealised diurnal variation signal with a maximum SST magnitude at approximately 15:00 local time. Satellite observations made by different sensors are indicated as blue dots which sample at different times of the day. Each measurement is slightly different depending on the local time of measurement and the phase of thermal stratification. The NGSST method distinguishes between a daily-mean SST that is based on a simple arithmetic mean of all available measurements and a daily minimum SST is defined as the lowest SST appearing just before local sunrise in a cloud-free and calm wind day.

The NGSST method applies a bias correction in which observed SST values are adjusted to account for diurnal stratification using the model of Kawai and Kawamura (2002). The NGSST is then derived based on an objective analysis described below.

WP-ID4.1.2 NGSST Objective analysis procedure

Objective analysis used to derive L4NGSST can be based on L2P input data. Assuming that the observations from satellites are available with irregular spatial and temporal gaps, the linear minimum mean square estimation of SST at the location (x,y) at time t is given by (Carter & Robinson, 1987),

$$\hat{\theta}(x,y,t) = CA^{-1}\phi \,, \qquad (4.1.1)$$

where $\hat{\theta}(x,y,t)$ is the estimated SST, ϕ is the matrix of SST observations, A^{-1} is the inverse of autocorrelation matrix between the observations, and C is cross correlation matrix between the estimations and observations of SST.

A simple analytical correlation function $C(r)$ was used (Carter & Robinson, 1987).

$$C(r) = (1 - r^2)\exp(-r^2/2) \quad (4.1.2)$$

$$r^2 = \left(\frac{\Delta x}{L}\right)^2 + \left(\frac{\Delta y}{L}\right)^2 + \left(\frac{\Delta t}{T}\right)^2 \quad (4.1.3)$$

Where Δx is zonal distance between the estimation and observation, Δy is meridional distance between the estimation and observation, Δt is temporal difference between the estimation and observation, L and T are spatial and temporal decorrelation scales. (In the current implementation, for AVHRR, all the SST data available in a 5x5 box (i.e. 0.05° latitude by longitude) were averaged.) Figure 4.1.3.1 shows the processing flow used to produce L4NGSST output.

WP-ID4.1.3 Summary of Japanese RDAC implementation of NGSST

In the regional implementation of NGSST at the Japanese RDAC, TMI, GMS and AVHRR data are currently used and an output grid size set to 0.05°. For TMI, a weight function was applied to the 3 pixels along zonal and meridional direction centred at the re-sampled pixel in order to correct for the sub-pixel effect. L and T were set as constant, i.e., 1° and 5 days. The previous-day and following-day data were used for the estimation of current-day SST. For each grid, firstly, quality control was conducted for the input data within 0.5° of the estimation. The criterion was set as,

$$\left| SST_i - \overline{SST_i} \right| \le 2K \quad (4.1.4)$$

Table 4.1.3.1 Specifications of satellite data used in the Japanese RDAC implementation of the NGSST

Satellite Sensor	Spatial Resolution	Temporal Resolution	Coverage	Accuracy
NOAA AVHRR	0.01°	Twice per day per satellite	20°N ~60°N 120°E~160°E	0.6 K
GMS S-VISSR	0.05°	Hourly	20°N ~60°N 120°E~160°E	0.8 K
TRMM MI	0.25°	Three days for full coverage	38°S ~38°N 0°E~360°E	0.7 K
TRMM VIRS	0.05°	Three days for full coverage	38°S ~38°N 0°E~360°E	0.7 K

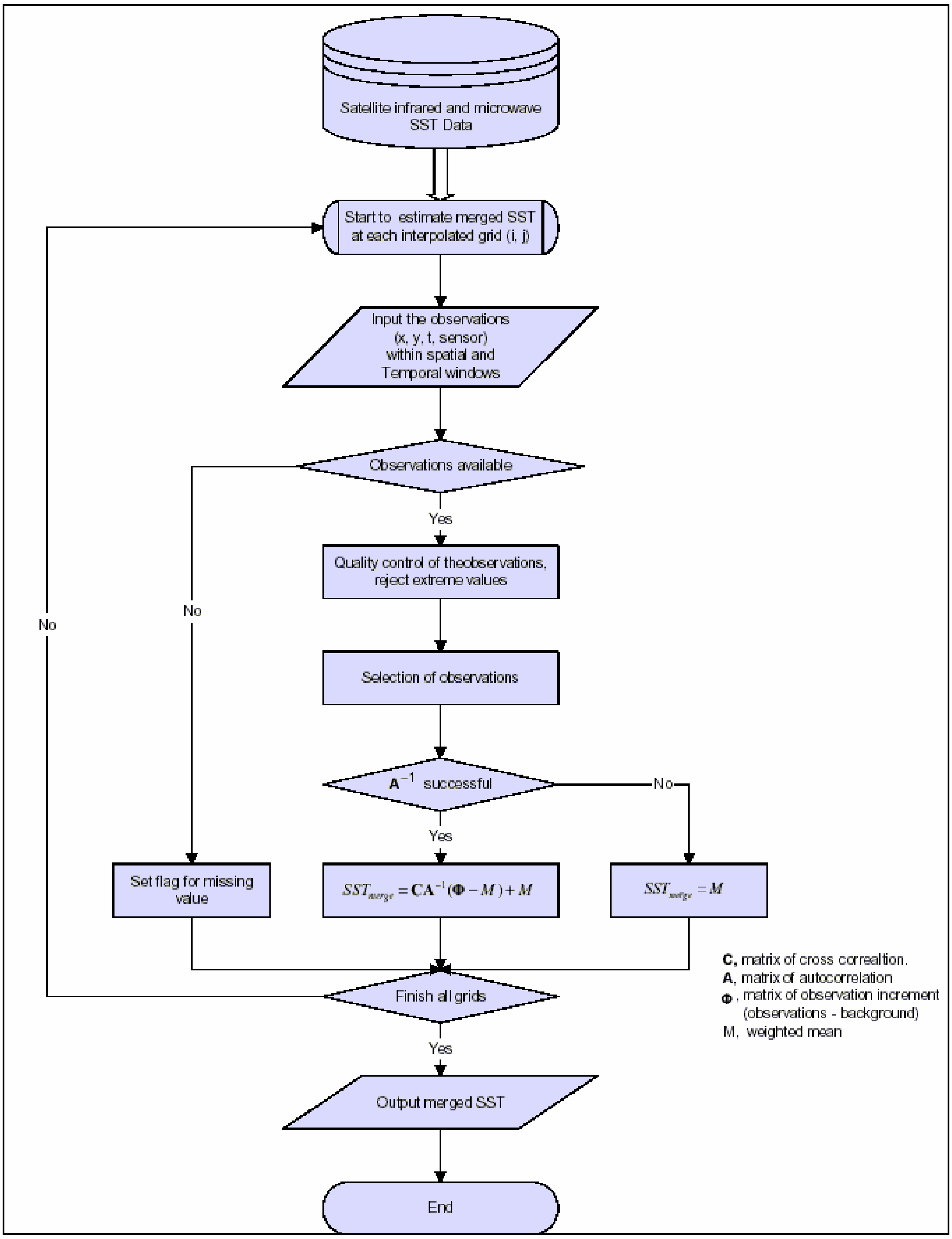


Figure 4.1.3.1 Flow diagram showing the NGSST analysis algorithm

Compared with the mean of the input SST values, data with the difference greater than 2 K were rejected. An *a priori* observation selection priority was set as AVHRR (VIRS), TMI and finally GMS, taking account of SST accuracy and spatial resolution. Only the highest priority SST was selected if several observations were available at the same location. The selected observations are then sorted by their cross correlation. The observation with the highest cross correlation is selected. Then the correlation of following observation with already selected data was compared with the correlation of the observations with the interpolated gird. If the former value is less than the latter value, the observation was selected, otherwise it is rejected. A threshold of the number was set to select the observations with the best cross correlation and in the current regional implementation, this threshold is set to 25. Hence, when many observations are available only the observations close to the interpolated grid are used. After these QC steps, a weighted mean is removed from the observations. The linear minimum mean square estimator was applied. The mean was added back after the estimation. If the inverse of the autocorrelation matrix was invalid, a weighed mean value was given for the output.

WP-ID4.1.4 Regional/seasonal dependent decorrelation scales for NGSSTv2.0

Using the cloud-free NGSST SST products, temporal/spatial variability of the SST fields are examined. The wavelet transformation (Hosoda and Kawamura, 2002a) was used to investigated the features of NGSST fields through comparisons with the buoy time-series SSTs. Monthly spatial decorrelation scales have been calculated by the wavelet analyses (Hosoda and Kawamura, 2002b) which show the temporal/spatial de-correlation scales are functions of season and region.

Based on this analysis, Equation (4.1.3) is now rewritten as,

$$r^2 = \left(\frac{\Delta x}{L(x,y,t)}\right)^2 + \left(\frac{\Delta y}{L(x,y,t)}\right)^2 + \left(\frac{\Delta t}{T(x,y,t)}\right)^2 \qquad (4.1.5)$$

The decorrelation scales, e-folding scales of SST variability, in the sea south of Japan (Figure 4.1.3.2) are order of 2-3 degree for space and order of 2-3 days for time, and are inhomogeneous, anisotropic and time-dependent. An example of the calculated decorrelation scales for March in this region is shown in Figure 4.1.3.6.

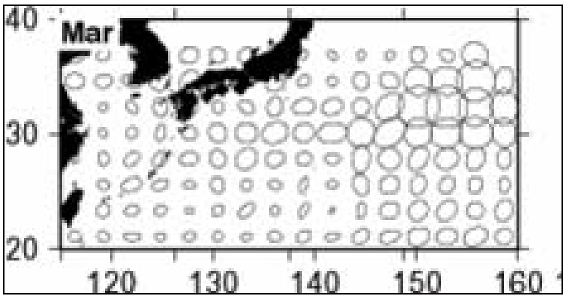


Figure 4.1.3.6 Decorrelation scales for the NGSSTv2 in her region of Japan.

WP-ID4.2 Compute L4 SSTfnd data products (L4SSTfnd)

The NGSSTv1 and V2 methodology described in WP-ID4.1 provides an estimate of the daily minimum SST1m once per day. Significant diurnal variation of SST1m could still be present in NGSST data streams. The purpose of defining an analysis procedure for L4SSTfnd data products is to provide a separate estimate of SST that is free of all diurnal effects (i.e., cool skin and warm layers) that is more representative of a mixed layer temperature.

WP-ID4.2.1 Bias correction strategy

Bias correction of all input data to an OI procedure is critical to obtaining a valid output (see for example Reynolds et al, 2002). As an initial bias correction scheme, the GDSv1 will use the L2P SSES derived bias (WP-ID3) prior to application of the OI procedure. In addition, bias due to diurnal stratification and cool skin effects may be accounted for using additional data derived using the **`diurnal_confidence`** procedure described in WP-ID2.2.

As the GHRSST-PP Science Team gains experience with the GDSv1, it is expected that the bias correction strategy will evolve and an upgrade of this component of the GDS is expected.

WP-ID4.2.2 The L4SSTfnd Optimal interpolation (OI) scheme

The Optimal interpolation scheme will follow that described by Reynolds et al. (2002). Appropriate decorrelation length scales will be specified by the GHRSST-PP Science Team for each sensor and region. All parameters associated with the operation of the L4SSTfnd OI scheme should be stored in a L4SSTfnd configuration file allowing for their easy editing and adjustment.

In parallel, it is expected that alternative methods to the scheme proposed by Reynolds et al (2002) will be explored and compared during the GHRSST-PP. for example, the fast OI scheme proposed by Menemenlis et al (1997) and Murray et al (

WP-ID4.3 L4 diurnal variation data fields

Diurnal variation data fields form part of each L4* data product and are designed to provide an estimate of diurnal variations (DV) in the SST signal. These data will enable users to estimate the magnitude and phase of DV by using an estimate of the SSTskin temperature at temporal intervals of 6 hours together with the SSTfnd (or SST1m). As an initial specification for the derivation of SSTskin data, the bias correction strategy described in Donlon et al (2002) may be used to bias adjust night-time Microwave and SSTdepth satellite observations to SSTskin. When available data from satellite sensors providing a direct estimate of the SSTskin (AATSR), MODIS as available) may be used directly. These data can then be used in an OI procedure to derive complete global fields. It is expected that as the GHRSST-PP matures, a more comprehensive strategy will be adopted to generate estimates of SSKskin from multiple sensors.

An estimate of the uncertainty of DV should accompany each estimate of peak warming. Additional flags can be used to indicate whether the warming was derived from direct observations (and their proximity in time to the peak warming), model based estimates (empirical or numerical), or if there was insufficient data. It is expected that these data fields will evolve as the GHRSST-PP Science Team gains more experience with the GDS system output.

WP-ID4.4 Generation of L4 netCDF and GRIB data files

Each L4* data set will be formatted and archived in netCDF or GRIB[8] format following the data format described in Appendix A1.2, A1.3 and A1.6

WP-ID4.5 L4 MMR-FR metadata records

In order for L4* data files to become "visible" to the global GDSv1 system and all other processing centres, a MMR-FR metadata record must be prepared and successfully

[8] The GRIB data format, while preferred by many operational meteorological agencies, will only be used if absolutely required. The standard GHRSST-PP netCDF data format is the preferred data format for all GHRSST-PP data products as this format provides the widest possible interoperability.

submitted to the GDS MMR by the responsible processing centre for all L4* data products.

WP-ID4.5.1 Format L4* MMR-FR metadata record

A MMR-FR metadata record should be prepared for each L4* data file according to the specification provided in Appendix A6.2.

WP-ID4.5.2 Register L4* MMR-FR metadata record with MMR system

L4* MMR-FR metadata records should be delivered to the MMR system as soon as possible after the L4 data set has been produced following the procedure described in Appendix A6.4.

WP-ID4.6 Extraction of L4* HR-DDS granules

HR-DDS data granules should be extracted from all L4* data products for each HR-DDS site defined in Appendix A5.5. HR-DDS data granules should be prepared according to the specifications provided in Appendix 5.

WP-ID4.7 Registration of L4* HR-DDS data granules using HR-DDS MMR-FR metadata

A HR-DDS MMR-FR metadata record must be prepared for all L4* HR-DDS granules. Each HR-DDS MMR-FR should be successfully submitted to the HR-DDS MMR by the responsible processing centre.

WP-ID4.7.1 Format L4* HR-DDS MMR-FR metadata record

A MMR-FR metadata record should be prepared for each L4* data granule according to the specification provided in Appendix A6.2.

WP-ID4.7.2 Register L4* HR-DDS MMR-FR metadata record at the HR-DDS MMR system.

L4* HR-DDS MMR-FR metadata records should be delivered to the HR-DDS MMR system as soon as possible after the L4* HR-DDS data granule has been produced following the procedure described in Appendix A6.4.

WP-ID4.8 L4* data processing status reports to OPLOG system

In order to maintain a record of L4 data processing, an entry within the GDS operational log should be made via e-mail as explained in Appendix 7 indicating the operational status of all L4 data processing activities.

WP-ID4.9 Extraction of L4* match up database file records (MDB-FR)

L4* MDB-FR should be extracted according to the procedures described in WP-ID3.

WP-ID5: Validation of GDS data streams

<table>
<tr><td colspan="3">WP-ID5 Validation of GDS data streams</td></tr>
<tr><td colspan="2">Work Package number :</td><td>WP-ID5</td></tr>
<tr><td colspan="2">Leader:</td><td>P. Minnett (pminnett@rrsl.rsmas.miami.edu)</td></tr>
<tr><td colspan="2">Aim:</td><td>To specify the procedures to validate the data streams used within the GDSv1</td></tr>
<tr><td>1</td><td colspan="2">Objectives:
1. To specify the procedures used for the validation of GHRSST-PP L2P and L4* data products.
2. To specify the diagnostic tests and tools that will be used to validate the GHRSST-PP methods using HR-DDS data granules and GHRSST-PP data sets.
3. To specify the reporting procedures for validation results to (a) the GHRSST-PP ISDI-TAG (b) the GHRSST-PP community and (c) data providers (if necessary).
4. To specify the feedback mechanisms between the GHRSST-PP GDS-v1 and the ISDI-TAG if problems are found and the likely procedure for action (a control loop)</td></tr>
<tr><td>2</td><td colspan="2">Description:
This section of the GDS-v1 describes in detail methods used to validate the data streams used within and generated by the GDSv1. Validation of the data products used with and generated by the GDSv1 provides the only independent method for assessing their accuracy and quality and is an integral component of the GDS.

WP-ID3 describes the process of matching independent in situ observations with satellite observations that are near contemporaneous. This WP is dedicated to the analysis of MBD data records and the wide validation of GHRSST-PP data products (L2P and L4*). The validation of satellite data is a complex process requiring careful attention to QC of all data sets. Consequently, a summary and outline specification for diagnostic tools and procedures is provided as a guide for the on-going validation and assessment of GHRSST-PP data products at RDAC and GDAC.

A set of reporting procedures are provided as guidelines to the GHRSST-PP validation effort that are designed to generate appropriate actions and solutions in case of problems as a control loop. In this way decisions relating to changes in the GDS can be properly documented and controlled.</td></tr>
<tr><td>3</td><td colspan="2">Inputs:
1. L2P and L4* in situ and satellite data
2. In situ observations
3. HR-DDS data granules</td></tr>
<tr><td>4</td><td colspan="2">Outputs:
1. MDB validation records
2. Diagnostic experiment reports and notifications
3. GHRSST-PP scientific and technical validation reports</td></tr>
<tr><td>5</td><td colspan="2">Acceptance tests:
1. L4* MDB-FR data records are ingested by the MDB
2. Diagnostic tools are functional
3. HR-DDS data granules are available and easily accessible</td></tr>
</table>

	4. Reporting procedures are functional 5. In situ and satellite data agree within expected limits based on sensor characteristics and uncertainties in retrieval algorithms. 6. Discrepancies between in situ and satellite data do not exhibit systematic dependencies on measurement (e.g. Satellite zenith angle) or environmental parameters. 7. Discrepancies between in situ and satellite data do not show significant trends with time. 8. Discrepancies between in situ and satellite data do not show dependencies on RDAC processing location.
6	**Metric for performance assessment:** 1. L4* MDB-FR are ingested at the MDB with a failure rate of < 2% 2. Appropriate action is taken in case of processing anomalies and adequate feedback is generated

Comparison with in situ measurements of SST is the primary mechanism for independently quantifying the accuracy of the satellite SST fields that are used and produced by the GDS. This requires that the input fields and the subsequently merged/analysed fields be compared with high-quality measurements of SSTskin and subsurface SST throughout the period of the GHRSST-PP. The comparisons need not, indeed in most cases cannot, be done in real time, but should be done in as short a time thereafter as practical (in the order of days). This is necessary to ensure that sensor problems are identified promptly and that the consequences of unanticipated geophysical events, such as volcanic eruptions, are quickly identified and their effects on the derived SSTs are quantified without delay. WP-ID3 describes the GHRSST-PP Match Up database (MDB) system that is used to store GHRSST-PP validation data records. This work package is dedicated to outlining the analysis of MDB data records for validation of GHRSST-PP data products.

The objectives of data product validation are several-fold. They include a monitoring of the accuracy of the on-board calibration procedures used by each of the sensors and the specification of regional and seasonal characteristics of the residual uncertainties in the algorithms used to correct for the effects of the intervening atmosphere (clear sky water vapour effects and cloud screening in the infrared; water vapour, surface wind speed and rainfall effects in the microwave). Validation should also identify errors consequent of the algorithms used to derive the merged and analysed SST fields.

Direct comparison between the separate satellite-derived SST fields is also a useful exercise to assess the relative accuracy of these and to confirm their compatibility before the merging operations. It provides a mechanism for assessing quickly the accuracy of SSTs derived from new sensors that become operational during the GHRSST-PP as a large number of cloud-free comparisons which span the global range of SST and atmospheric variability can be generated in a relatively short time., It also provides a link between observations at high-accuracy but narrow swath (e.g. AATSR) or low resolution (e.g. AIRS) and the independent validating measurements through wide-swath, high resolution sensors (e.g. AVHRR, MODIS) which can act as a transfer standard. Otherwise, achieving the number of coincident measurements between these sensors and the in situ data that are necessary to define the uncertainties through a wide sample of parameter space may take too long.

The individual SST fields that require validation are presented in Appendix 3.2, and the in situ measurements that will be used for the validation include well-calibrated radiometers and interferometers mounted on ships and aircraft, subsurface measurements taken from moored and drifting buoys, the latter including surface drifters and autonomous profilers. These are presented in Appendix 3.4. Auxiliary supporting satellite data streams that can be used to assist in the validation exercise are listed in Appendix 3.3

Over the duration of the GHRSST-PP, the availability of validating sensors is likely to be those that are available now, and which are being used to validate the current generations of satellite sensors. New shipboard and aircraft sensors may become available in the duration of the GHRSST-PP, but confidence in the accuracy of these sensors must first be established before their data are incorporated into the GHRSST-PP. One mechanism for this would be through collocated deployments with established sensors on the same ship or aircraft.

The instruments used to provide the validation data must themselves be subject to accurate and continuing calibration, traceable to national standards, and be deployed according to the appropriate guidelines and protocols.

Because of the different volumes of data involved, it is more efficient for the in situ measurements to be provided to the RDACS, instead of the satellite data being delivered to the supplier of the validating measurements. The routinely-collected validation data could be delivered by periodic ftp-pull to the RDACS, while the episodic data could be delivered by ftp-push or by ftp-pull after special notification.

WP-ID5.2 The procedures and rules that will be used to validate GHRSST-PP data products.

The procedures to be adopted for the validation of the GHRSST-PP SST fields are those required to fill the elements in the MDB FR validation Records (described in Appendix A4.1) and to conduct the comparisons between the satellite-derived SST measurements and the reference data. The information in the validation data sets should be determined and reported in terms of the user requirements. These include not only absolute uncertainties but also relationships and systematic dependences between these errors and other geophysical variables that are represented in the auxiliary data sets.

The following rules are specified:

- **Rule 5.2a:** The validation will be expressed in terms of the statistics of the relationship between the measured satellite SST field and the reference values (measured – reference; i.e. an error, not a correction).
- **Rule 5.2b:** As a minimum the error will be expressed as a mean and standard deviation. But, wherever possible, dependences of the errors on time, geographical location, and the auxiliary variables, such as surface wind, water vapour, aerosols, etc, should be explored and reported.

It is expected that significant scientific study, on a case by case basis, will be required to derive accurate validation reports using MDB data records. Accordingly, GHRSST-PP scientific and technical reports will form the main output of the validation activities.

WP-ID5.3 Diagnostic tests based on the GHRSST-PP high resolution diagnostic data set (HR-DDS)

The GHRSST-PP high resolution diagnostic data set (HR-DDS) provides an extremely rich source of information that can be used to conduct diagnostic experiments on GHRSST-PP data products and L2P data streams. After each satellite product has been calibrated and quality controlled as accurately as possible, several diagnostic products will be computed to assure that the input satellite data are as accurate as possible. This is necessary because the errors change with time, e.g., orbits decay or may be modified due to mission requirements, satellite instruments age and are replaced by new instruments. These diagnostics will be performed on small time and space scales within the HR-DDS system. However, some of the comparisons that are required involve in situ data which are sparse and only helpful on relatively large time and space scales. In addition, it is necessary at this stage to develop a relatively small set of diagnostics so that they are easier to monitor.

WP-ID5.3.1 Additional Tools

In addition to the in situ SSTs produced from drifting and moored buoys, it is planned to use two additional analysis products. The first is the optimum interpolation version 2 (OI.v2) analysis produced by a blend of AVHRR and in situ data (ship and buoy). The analysis is produced operational weekly on a one-degree spatial grid and is described in Reynolds, et al. (2002). The second is the monthly climatology produced from OI.v2 and other analyses with a base period of 1971-2000. The analysis method is described in (Smith and Reynolds, 1998). Xue, et al. (2003) have updated the base period using the Oi.v2 and describe the changes in the base period with time. Both data sets are available in real-time at
http://www.emc.ncep.noaa.gov/research/cmb/sst_analysis/
and
http://www.cpc.noaa.gov/products/predictions/30day/SSTs/sst_clim.html, respectively. These two products are diagnostic tools only and are not included in any final GHRSST product.

WP-ID5.3.2 Time series diagnostics

Within the GDSv1, the following time series will be computed and posted to the GHRSST web page under a diagnostic results web page.

1. A weekly time series will be produced of the collocated mean and RMS difference of each satellite product with respect to drifting and moored buoys. This will be done separately for day and night and for 4 regions: global, between 30N and 30S, south of 30S and north of 30N.

2. Because the differences in the above time series are sampled at buoy locations, they are not sampled uniformly. Thus an additional set of weekly time series will be produced of the average SST anomaly for each satellite product and for the Oi.v2 analysis. The anomaly will be computed using the OI.v2 climatology interpolated spatially and temporally as required. This will again be done for the 4 regions, above.

3. To monitor differences on shorter time scales, a daily time series will be produced for each satellite product showing the mean and standard deviation of the anomaly. This will be done separately for day and night and be done for 4 regions listed above.

WP-ID5.3.3 Other diagnostics

These initial diagnostics are produced to identify possible problems. Further diagnostic products will be computed as required. The additional products may include regional time series and mean differences between products.

WP-ID5.3.4 Additional notes on satellite operations

As satellites change and replace or upgraded with time, changes in the quality of the data may be impacted. The GHRSST web page will include notes for users for each satellite product. These notes will include but not be restricted to any change in the instrument or orbit that may impact the data.

WP-ID5.4 The procedures and rules that will be used to validate GHRSST-PP data products external the DDS system.

It may be necessary to include particularly rich or promising validation data sets that lie outside of the DDS system. These can be accommodated by focussed validation exercises to make use of these sources, but operational constraints may require that this be done in a 'reprocessing' context. If a new geographically well-defined source of validation measurements becomes available, with a reasonable prospect of continuing operation over a significant time interval, then that location can be included in a revised DDS list. This revision will be approved by the "Problem Resolution Board" (Appendix 8) taking into account the likely benefits and the logistical costs involved.

WP-ID5.5 To specify the reporting procedures for validation results to (a) the GHRSST-PP ISDI-TAG (b) the GHRSST-PP community and (c) data providers (if necessary).

The results of the validation and quality assurance exercises will be made available to all interested parties through a series of web pages that graphically portray the findings. ASCII files of the validation records will be available for ftp-get through the same set of pages. An example such a system, the MODIS Oceans QA Browse Imagery system (MQABI) is accessible at http://jeager.gsfc.nasa.gov/browsetool/ and could serve as a model for the GHRSST-PP scheme. After tailoring to meet the requirements specific to GHRSST, it would be installed at all RDACS and the GDAC.

It is anticipated that the GHRSST-PP will be able to use the MODIS (Ocean) Quality Assurance Browse Imagery (MQABI) code-base (http://mquabi.gsfc.nasa.gov) that allows comprehensive browsing and download of GHRSST-PP HR-DDS granules.

WP-ID5.6 To specify the feedback mechanisms between the GHRSST-PP and the ISDI-TAG if problems are found and the likely procedure for action (a control loop)

The effort invested in the validation exercise is wasted unless the information gained can be used to refine the quality flags and improve the products developed by the GDSv1. This requires a monitoring activity of all of the fields being validated. A 'Monitor' should be appointed to analyse the content of the validation and QA web pages on a daily basis and generate brief reports on a weekly basis. These will be distributed to the GHRSST-PP Science Team by e-mail and posted on the web pages (possibly with initial password protections for Science Team access only). The web pages should allow users to report problems to the 'Monitor' who should respond directly to the user and include the contents in the weekly reports. A sub-set of the Science Team, including the Chairman, will be appointed to a "Cal-Val Problem Resolution Board" that should meet by conference call on a monthly basis to resolve problems (by assigning the issue to the most competent Science Team Member as necessary), discuss trends and try to anticipate problems and their solutions. They will also recommend to the Science Team changes to algorithms and processing procedures that would lead to improved products. When an anomalous situation arises, such as a sensor failure or a volcanic eruption, an alarm will be raised by the Monitor by email to the "Cal-Val Problem Resolution Board" which should convene a conference call at the earliest opportunity.

The following rules are specified:

- **Rule 5.6a:** Weekly reports of all cal-val activities should be generated throughout the operational phase of the GHRSST-PP. These should be sent to the GHRSST-PP Science Team.
- **Rule 5.6b:** Monthly conference calls should be convened by the Chairman of the Cal-Val Problem Resolution board (CVPRB).
- **Rule 5.6c:** If a Cal-Val problem is found the CVPRB should convene a dedicated conference call at the earliest possible opportunity in order to establish an appropriate course of action.

WP-ID-6 Commissioning and operating the GDS

WP-ID6	**Commissioning and operating the GDS**	
Work Package number :	**WP-ID6**	
Leader:	**[TBD]**	
Aim	**To specify the acceptance tests and metrics that will be used to commission and monitor the GDS-v1.**	
1	**Objectives:** 1. To describe the GDSv1 reference processor. 2. To define the GDSv1 reference data set. 3. To specify the acceptance tests that will be used to commission the GDS-v1. 4. To specify the metrics that will be used to monitor the performance of the GDS-v1. 5. To specify the criteria that will be used to raise a significant processing model failure (SPMF) of the GDS-v1. 6. To specify the criteria that will be used to raise a significant processing model anomaly (SPMA) of the GDS-v1 7. To specify the procedures that will be implemented in the case of a SPMA or SPMF to reinstate data processing as soon as possible.	
2	**Description:** The GDS-v1 will be commissioned to run at RDAC and GDAC facilities within the GHRSST-PP. It is foreseen as a demonstration system having an operational capability. This WP specifies the criteria that will be used to commission GDS-v1 at each RDAC and GDAC. A reference test data set (TDS) and a reference processor (RP) will be used to verify that each RDAC and GDAC as implemented the GDSv1 in an identical manner. This WP also specifies the procedures that should be followed in the case of a significant processing model failure (SPMF) or significant processing model anomaly (SPMA).	
3	**Inputs:**	
4	**Outputs:** 1. GDSv1 test data set 2. GDSv1 reference processor 3. Commissioning and operating manual	
5	**Acceptance tests:** [TBD]	
6	**Metric for performance assessment:** [TBD]	

This part of the GDS-v1 describes how the GDSv1 will be commissioned at each RDAC and GDAC using a reference processor system. The reference processor will be a scientific (as opposed to an operational) implementation of the GDSv1. The reference processor will be used together with a test data set (TDS) to verify that RDAC and GDAC implementations of the GDSv1 are capable of providing identical output.

In addition, the reference processor can be used to assess the implications of proposed GDS upgrades prior to operational implementation.

WP-ID6.1 The GDSv1 reference data processor

The GDSv1 reference data processor should be a scientific implementation of the GDSv1. A fourth generation programming language (e.g., IDL or MatLab) should be used to implement the reference processor so that visualisation of data variables is relatively straightforward at any point in the processing system. The reference processor should consider the following aspects:

1. The reference processor should ensure that as many (if not all) of the local data processing steps followed at each RDAC specific to regional data streams are included.
2. The reference processor should be made available to RDAC and GDAC as a reference tool that can be used to troubleshoot regional problems.
3. A master version of the reference processor should be implemented and maintained as such at a GDAC centre.
4. the reference processor should include

    (a) All L2 to L2P conversions
    (b) L4 derivations
    (c) An example MMR-FR metadata reference generator
    (d) An example HR DDS granule extractor
    (e) An example HR-DDS MMR-FR metadata generator
    (f) An example SSES generator

It is foreseen that RDAC and GDAC centres will provide example code for the implementation of regionally specific I/O and data analysis modules and that the reference processor code base will be "open source".

WP-ID6.2 GDSv1 Test Data Set (TDS)

The GDS is a shared system and must be implemented at several different RDAC/DGDAC. If the GHRSST-PP regional task sharing approach is to be successful and global data products produced from independent data processing centres, it is critical that each RDAC and GDAC produce the same quality and standard of data products. The purpose of a test data set is to implement the GDSv1 in a non-operational environment and provide a common data set to test the implementation of the GDSv1 at each RDAC/GDAC. The GDSv1 test data set (TDS) will be defined and processed by the reference processor to provide a reference output data set. Each RDAC and GDAC may then use the TDS to produce an output data set that can be compared to the reference processor output. Differences between each RDAC/GDAC system can be readily identified by comparing the output and problems solved using the TDS.

The GDSv1 reference data set should include example L2 data streams from all RDAC and GDAC in the native format expected at RDAC and GDAC. Each L2 data stream should be extracted for a reference data period which is defined as July-August 2002. A master copy of the GDSv1 reference data set should be made available as a DVD-ROM and distributed to all RDAC and GDAC centres.

The TDS will be available to all participants within the GHRSST-PP. The TDS will include the following:

- **Part 1:** 2 days of L2 data files having global coverage from each L2 satellite data stream considered by the GHRSST-PP as defined in [RD-3, Appendix 3]. Both polar orbiting and geostationary observations will be included.
- **Part 2:** 1 week of in situ data straddling the same period as the data sets described in Part 1.
- **Part 3:** An example GHRSST-PP Match-Up Database[9] with entries for each sensor [RD-3].
- **Part 4:** Example GHRSST-PP netCDF L2P, SSTfnd, HR-DDS granule format data files and associated metadata records.

WP-ID6.3 Acceptance tests to commission the GDS-v1 at each RDAC and GDAC

In addition to the reference processor, a set of basic functional tests should be defined that can be used to ensure that the regional task sharing processing system defined by RDAC and GDAC interactions and data exchange is functional. The aim of these tests is to ensure that the GDS system is working as an integrated system. Metrics must be agreed and assigned by the GHRSST-PP Science Team, RDAC and GDAC teams according to known priorities and limitations

As the GHRSST-PP is a real time operational system it is imperative that during "operational" phases of the project data are made available and delivered in a timely manner. The criteria for the exchange and delivery between GDAC and RDAC and from data provider to the RDAC have differing priorities that are based on a need to provide sufficient time to produce data products. The following GDSv1 rules are specified:

- **Rule 6.3a:** Each L2 satellite data stream available for a given processing window should be processed to L2P products as soon as possible after delivery to a processing centre. In any event, L2P should be complete within 3 hours of the end of a PW time period in order to allow sufficient time for the generation and delivery of L4 data products.
- **Rule 6.3b:** L4 data products should be produced within 6 hours following the end of an analysed product processing window (APPW). After this time, data produced in the following APPW may be more appropriate.
- **Rule 6.3c:** All metadata records should be automatically delivered to the MMR immediately following successful data product generation. In any case, metadata should be delivered to the MMR no later than 60 minutes following these events.
- **Rule 6.3d:** All data extracted for use within the HR-DDS system should be ingested into local OpENDAP systems and made available to the GHRSST-PP community within 3 hours following extraction via a Live Access Server (LAS) style interface. Direct ftp access should also be provided. This is important in order that other RDAC/GDAC systems may take advantage of the ability to monitor the quality of GHRSST-PP data streams in real time.

[9] TDS MDR records may be simulated in order to provide functionality.

WP-ID6.4 To specify the criteria that will be used to raise a significant processing model failure (SPMF) or a significant processing model anomaly (SPMA) for the GDS-v1

A significant Processing Model Failure (SPMF) is the term used to refer to the situation where an RDAC or GDAC has failed to provide an expected output data stream within the framework of the GDSv1. For example, an SPMF will be raised if RDAC data products are delivered too late for inclusion into the appropriate PW or APPW. A significant processing model anomaly refers to a failure of the GDS that does not result in a SPMF but was of sufficient importance to warrant further investigation.

A list of expected SPMF situations should be identified by RDAC and GDAC teams. As a guideline the following criteria should be used to establish potential SPMF:

- What are the significant weaknesses in the regional implementation of the GDS?
- Where are the likely failures expected to occur?
- What should the action be at RDAC to address these failures should they occur

WP-ID6.5 To specify the procedures that will be implemented in the case of a SPMA or SPMF to reinstate data processing

In the case of a SPMA or a SPMF the GDS TAG Problem resolution Board (PRB) will be notified and take appropriate action as described in Appendix A8.

WP-ID7: Upgrade paths from GDS-v1 to GDS-v2

WP-ID7	Upgrade paths from GDS-v1 to GDS-v2
Work Package number: **Leader:** **Aim**	**WP-ID7** **Craig Donlon (craig.donlon@jrc.it)** **To identify the components of the that will be upgraded from the GDS-v1 to the GDS-v2**
1	**Objectives:** 1. To identify and document the components of the GDS-v1 that require significant upgrade
2	**Description:** As the GHRSST-PP progresses it is clear that significant upgrades to the GDS will be required as more experience is gained during its implementation. This WP is focussed on establishing which components of the GDS will require upgrade and modification to GDSv2.
3	**Inputs:** 1. GDSv1 specification (this document)
4	**Outputs** 1. Within 6 months of GDSv1 commissioning an assessment of upgrade requirements should be complete and reported 2. Within 9 months of GDSv1 commissioning end, a detailed plan for the upgrade of the GDSv1 system should be available.
5	**Acceptance tests:** N/A
6	**Metric for performance assessment**: N/A

The following sections highlight the major components of the GDSbv1 processing system that will be upgraded.

WP-ID7.1 Changes in L2 inputs

It is foreseen that L2 data streams will change throughout the life of the GHRSST-PP. The GDS must be capable of bringing L2 data streams in and out of the processing system without affecting other components of the system.

WP-ID7.2 Upgrade of IPCV and MPCF procedures

The GDSv1 procedures for the derivation of IPCV and MPCV values described in GDS WP-ID2.1.1.14 are quite basic. It is clear that significant improvements can be made leading to a better quality of pixel error statistics. It is foreseen that this component of the GDSv1 will be upgraded following further research and development within the GHRSST-PP Diagnostic Data Set.

WP-ID7.3 Changes to L2P quality control procedures

It is foreseen that minor modifications to the procedures used to QC L2 input data and derive L2P data files may be required once significant experience is gained with the GDSv1.

WP-ID7.4 Changes to analysis procedures and data products

The GDSv1 SST1m analysis procedure is based on the New Generation SST v2 scheme and is considered a first step towards an operationally efficient analysis scheme that will deliver estimates of the subsurface SST in real time. It is expected that as the GHRSST-PP project develops significant modifications to this methodology will be made in terms of the bias adjustment strategy.

The methodology outlined for the generation of L4SSTfnd is basic in the GDSv1 in order to have an operational system in place. It is foreseen that a significant upgrade of this scheme will be undertaken in the GDSv2.

WP-ID7.5 Changes to data delivery and data format

As experience is gained in the distribution and application of GDSv1 data products, data product format changes are foreseen based on user requirements and data delivery constraints.

WP-ID7.6 GDSv1 reference processor upgrade

The GDSv1 reference processor system should be used to test all proposed upgrades so that the implications of the proposed changes may be evaluated prior to operational implementation.

WP-ID7.7 Upgrade of L2P confidence record `DV_confidence` and `DV_shape` variable derivation

An upgrade of all diurnal variation parameters and derivations is foreseen in the GDS v2 based on further R&D within the GHRSST-PP.

References

Brisson, A., S. Eastwood, P. Le Borgne and A. Marsouin, O&SI SAF Sea surface Temperatures: pre-operational results, Proceedings of the 2001 Meteorological Satellite Data User's conference, Antalya, Turkey, 2001.

Carter, E. F. and A. R. Robinson, 1987: Analysis models for the estimation of oceanic fields. J. Atmos. Oceanic. Technol. , 4, 49-74.

Casey, K., et al. the GHRSST-PP Reanalysis Project Strategy, (in prep) available form the GHRSST-PP International Project Office, 2003.

Cavalieri, D. J., J. Crawford, M. Drinkwater, W. J. Emery, D. T. Eppler, L. D. Farmer, M. Goodberlet, R. Jentz, A. Milman, C. Morris, R. Onstott, A. Schweiger, R. Shuchman, K. Steffen, C. T. Swift, C. Wackerman, and R. L. Weaver., NASA sea ice validation program for the DMSP SSM/I: Final report. NASA Technical Memorandum 104559, 1992.

Cavalieri, D. and J. Comiso. Algorithm Theoretical Basis Document for the AMSR-E Sea Ice Algorithm, Revised December 1. Landover, MD, USA: Goddard Space Flight Center, 2000. http://nsidc.org/data/docs/daac/ae_si12_12km_tb_sea_ice_and_snow.gd.html and http://polynya.gsfc.nasa.gov/seaice_projects.html.

Donlon, C. J., T. J. Nightingale, T. Sheasby, J. Turner, I. S. Robinson and W. J .Emery, The Thermal Skin Temperature of the Ocean at High Wind Speed, Geophys. Res. Lett., Vol. 26, No. 16, 2505-2508, 1999.

Donlon, C. J., P. Minnett, C. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward and, J. Murray, Towards Improved Validation of Satellite Sea Surface Skin Temperature Measurements for Climate Research, J. Climate, Vol. 15, No. 4, 353-369, 2002.

Donlon,. C. J., The GODAE High Resolution Sea Surface Temperature Pilot Project: Strategy and Initial Implementation Plan, available form the GHRSST-PP International Project Office, 2002a.

Donlon, C. J., Proceedings of the Second GODAE High Resolution Sea Surface Temperature Pilot Project Workshop, Tokyo, Japan, May 13-15th 2002, available form the GHRSST-PP International Project Office, 2002b.

Donlon, C. J., Proceedings of the Third GODAE High Resolution Sea Surface Temperature Pilot Project Workshop, ESA/ESRIN, Frascati, Italy, December 2-4, 2002, available form the GHRSST-PP International Project Office, 2003a.

Donlon, C. J., The GODAE High Resolution Sea Surface Temperature Pilot Project Development and Implementation Plan (GDIP), available form the GHRSST-PP International Project Office, 2003b.

Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson and G. S. Young, Cool-skin and warm-layer effects on sea surface temperature, J. Geophys. Res., 101, 1295-1308, 1996.

Fieguth P.W., Allen M.R., Murray M.J., Hierarchical Methods for Global-Scale Estimation Problems, Canadian Conf. on Elec. Eng., Waterloo, Canada,161-164, 1998.

Fieguth, Khellah, Murray and Allen, Data fusion of SST data,Proc. IEEE 2000, 5, 2111-13, 2000.

Faugere,Y., P.Le Borgne and H.Roquet, Realisation d'une climatologie mondiale de la temperature de surface de la mer a echelle fine, La Meteorologie, 35, 24-35, 2001.

Gardashov, R. and M. C. Barla, the calculation of the distribution of the Sun glitter radiance on the ocean surface by observing from a geostationary orbit, Imt. J. Rem. Sens., 22, No. 15, 2939-2952, 2001.

Gentemann, C. L., C. J. Donlon, A. Stuart-Menteth and F. J. Wentz, Diurnal signals in satellite sea surface temperature measurements, in press, Geophysical Research Letters, 2003.

Guan, L. and H. Kawamura (2002): Merging satellite infrared and microwave SSTs—Methodology and Evaluation of the new SST, Submitted to J. Oceanogr. (Under review)

Guan Lei, H. Kawamura (2003): SST availabilities of satellite infrared and microwave measurements, J. Oceanogr., 59, 201-209.

Hosoda, K. and H. Kawamura (2002a): Examination of the merged sea surface temperature using wavelet analysis, submitted to J. Oceanogr. (Under review)

Hosoda, K. and H. Kawamura (2002b): Time-dependent spatial decorrelation scales of sea surface temperature field derived from the merged sea surface temperature, submitted to J. Oceanogr. (Under review)

Josey, S. A., E. C. Kent and P. K. Taylor, The Southampton Oceanography Centre (SOC) Ocean - Atmosphere Heat, Momentum and Freshwater Flux Atlas. Southampton Oceanography Centre Report No. 6, 30 pp. plus figs, 1998.

Kawai, Y. and H. Kawamura, Evaluation of the Diurnal Warming of Sea Surface Temperature Using Satellite-Derived Marine Meteorological Data, Journal of Oceanography, Vol. 58 (No. 6), pp. 805-814, 2002

Kawai, Y. and H. Kawamura (2002): Evaluation of the diurnal warming of sea surface temperature using satellite-derived meteorological data, J. Oceanography, 58, 805-814.

Menemenlis, D., P. Fieguth, C. Wunch and A. Willsky, Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data, J. Geophys. Res.,102, 10,573-10,584, 1997.

Murray M.J., Assimilation Techniques for blending GOES & POES SST retrievals., RALTechnical Report RAL-TR-2002-017, 2002.

Murray, M.J., Allen, M.R., Mutlow, C.T., and Llewellyn-Jones, D.T. Investigating tropical intraseasonal variability in sea surface temperature using data from the Along Track Scanning Radiometer. Proc. of the 2nd ERS-1 Symposium, Hamburg 1993, ESA SP-361, 529-533, 1994

Pinnock S. and C. J Donlon, Implementation of the GHRSST-PP High Resolution Diagnostic Data Set, GHRSST-PP Document reference GHRSST/14, available form the GHRSST-PP International Project Office, 2003.

Podesta, G and Evans, R, AVHRR Pathfinder Oceans Matchup database 1985-1997 (version 19.0), see http://www.rsmas.miami.edu/groups/rrsl/pathfinder/Matchups/match_index.html, 2003.

Reynolds, R. W., A real-time global sea surface temperature analysis, J. Climate, 1, 75-86, 1988

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes and W. Wang, An improved in situ and satellite SST analysis for climate. J. Climate, 15, 1609-1625, 2002.

Rossow, W. and L. Gardner, Selection of a map grid for data analysis and archival, Journal of Climate and Applied Meteorology, 23, 1253-1257, 1984.

Smith, T. M. and R. W. Reynolds, 1998: A high resolution global sea surface temperature climatology for the 1961-90 base period. J. Climate, 11, 3320-3323.

Tanahashi, S., H. Kawamura et al. (2002): Diurnal Variations of Sea Surface Temperature over the Wide-ranging Ocean Using VISSR on board GMS, J.Geophys. Res., (in press)

Wick, G. A. The First Report of the GHRSST-PP In situ and Satellite Data Integration-Technical Advisory Group (ISDI-TAG), GHRSST-PP Document reference GHRSST/13 v1.0, available form the GHRSST-PP International Project Office, 2002.

Xue, Y., T, M. Smith, and R. W. Reynolds, 2003: Interdecadal Changes of 30-Year SST Normals during 1871-2000. J. Climate, 12 (in press).

Appendix I. GDSv1 data product format specification

This appendix provides a detailed technical description of GDSv1 output data products. In its initial configuration, the GDS-v1 will output four SST data products according to the specifications laid out in Table A1.

Table A1 Data product families produced by the GDSv1. The reference section indicates the reference section in this Appendix

	L2 Pre-Processed	Collated SST	Merged SST	Analysed SST	Diurnal Variation
Acronym	L2P	L3C	L3M	L4	L4DV
Description	Native L2 data streams reformatted into netCDF	Single sensor L2P data that have been re sampled onto the GDSv1 output grid defined in Table A1.1	Multiple sensor L3C data that have been merged together to provide a complete SST field.	L4 analysed data (NGSST and SSTfnd) provide an estimate of the SSTfnd that is free of diurnal variability.	L4DV data products provide an estimate of diurnal variation for each grid cell of a L4A data product
Grid specification	Native to L2 data format	Table A1.1	Table A1.1	Table A1.1	Table A1.1
Temporal resolution	Native to data stream	Native to data stream: Consolidated for each GDS processing window	6 hourly PW centred at: 00:00, 06:00, 12:00 and 18:00 UTC	Analysed product processing window	12 hourly Analysed product processing window
Delivery timescale	As available	Single file data as available. Consolidated data products within 3 hours of a given processing window	Within 3 hours of a given processing window	Within 6 hours of an APPW	Within 6 hours of an APPW
Target accuracy	Native to data stream	Native to data stream, verified by QC procedures	< 0.5 K absolute 0.1 K relative	< 0.4 K absolute) 0.1 K relative	1K
Error statistics	Native to data stream if available	rms. and bias for each data stream at every grid point	rms. and bias for each input data stream at every grid point	rms. and bias for each output grid point (no input data statistics are retained)	Not provided
Coverage	Native to data stream	Regional and global	Regional and Global	Global	Global
Confidence data	Table A3.2.2	Table A3.3.2	Table A3.3.2	Table A3.4.2	Table A3.5.2
Nominal	netCDF see Section	netCDF – Appendix A2.1	netCDF – Appendix A2.1	netCDF – Appendix	netCDF – Appendix

data format	A2.1	GRIB - Appendix A2.3	GRIB - Appendix A2.3	A2.1 GRIB - Appendix A2.3	A2.1GRIB - Appendix A2.3

A1.1 GDSv1 operational output grid specifications

Table A1.1.1 describes the GDSv1 preliminary output grid specification that is applicable to L4* data products produced by the GDSv1.

Table A1.1 GHRSST-PP global coverage operational output grid specification

Description	
Boundary	90S- 90N and 0°-360°W
Projection	Equidistant Cylindrical
Reference Ellipsoid	WGS84
Grid cell size	9.28km (1/12°) at the equator

A1.2 GDSv1 netCDF common data product file structure

The GDSv1 netCDF file format is designed to be compatible with the GHRSST-PP Diagnostic Data Set (DDS) and the GODAE Data and Product Server Project. GDSv1 netCDF files are based on the attribute data tags defined by the Cooperative Ocean/Atmosphere Research Data Service (COARDS) and climate and forecast (CF) metadata conventions. The CF convention generalise and extend the COARDS convention but relaxes the COARDS constraints on dimension order and specifies methods for reducing the size of datasets. The purpose of the CF conventions is to require conforming datasets to contain sufficient metadata that they are self-describing in the sense that each variable in the file has an associated description of what it represents, including physical units if appropriate, and that each value can be located in space (relative to earth-based coordinates) and time (NetCDF Climate and Forecast (CF) Metadata Conventions, Version 1.0-beta5, 2003).

The following web reference documents are applicable to GDSv1 netCDF file formats:

netCDF:	http://www.unidata.ucar.edu/packages/netcdf/
CF conventions:	http://www.cgd.ucar.edu/cms/eaton/cf-metadata/CF-working.html
COARDS conventions:	http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html
UDUNITS:	http://www.unidata.ucar.edu/packages/udunits/

GDSv1 data files have a netCDF specification that is configured with four common variables as shown in Figure A1.1.

Global attributes	Mandatory
[n x m] array of SST data (e.g., SST, wind speed etc)	Mandatory
geolocation data (specific to Gridded data set)	Mandatory
[n x m] array of confidence flags (specific to variable data set)	Mandatory

Figure A1.1 Schematic representation of an GDSv1 netCDF output file identifying the major file components.

Table A1.5 describes the structure of an example GDSv1 L2C SST data file. Note that in the context of netCDF, a *variable* refers to data stored in the file as a vector or as a multidimensional array. Global attributes (rows 1-17) are used to hold information which applies to the whole file, such as the data set title. The global attributes shown in the example are the minimum set required for a valid GDSv1 netCDF file. Each individual variable can also have its own attributes.

In this example, the variable CSST consists of a 2-D array [n x m] of sea-surface temperature measurements. In order to save disk space, the 32-bit floating point temperature measurements have been scaled onto 16-bit short integers using the **add_offset** and **scale_factor** attributes. The floating point CSST values can be recovered using:

$$CSST_{float} = (\,\mathbf{scale\_factor} \times CSST_{short}\,) + \mathbf{add\_offset}$$

Also associated with the CSST variable are attributes describing the units, a longer descriptive version of the variable name, and a fill value. S.I units should be used and described by a character string which is compatible with the Unidata UDUNITS package. The CSST variable fill value is used to indicate array elements that do not contain a valid measurement.

As the GHRSST-PP output grid (defined in Appendix A1 Table 1.1) is fixed, two variables are included which provide the scales for the latitude and longitude axes of the CSST array, these are the so-called *coordinate variables*. These coordinate variables have not been transformed from floating point to integers, since they take up only a small fraction of the total required disk space. The only attributes defined in this case are the units. No fill value attribute is defined because vector coordinate variables should never have missing values.

A series of confidence data flags are supplied in a 2 dimensional [n x m] array confidence variable. These data are specific to each scientific data type (L2P and L4*).

Note that it is relatively easy to extend the format of a netCDF data file as another variable data set may be appended to the end of the file without modifying the original file format structure.

Table A1.5 An example GDSv1 netCDF data file.

Variable		Attribute	Type	Notes
1		**Filename**	Char	*e.g.* EUR- ATS_MET_2P-20030621T18-V1.0
2		**entry_id**	Char	Same as metadata field **Entry_ID** (see §4-6).
3		**Location**	Char	GHRSST-PP data centre code at which the data Set was created
4		**creation_date**	char: "yyyy-mm-dd"	UTC *e.g.* "2002-10-16"
5		**start_date**	char: "yyyy-mm-dd"	UTC, *e.g.* "1999-07-01"
6		**start_time**	char: "hh:mm:ss"	UTC, *e.g.* "14:12:59"
7		**stop_date**	char: "yyyy-mm-dd"	UTC, *e.g.* "1999-07-01"
8		**stop_time**	char: "hh:mm:ss"	UTC, *e.g.* "14:13:02"
9		**southernmost_latitude**	Float	*e.g.* 34.0
10		**northernmost_latitude**	Float	*e.g.* 36.0
11		**Westernmost_longitude**	Float	*e.g.* -16.0
12		**Easternmost_longitude**	Float	*e.g.* -14.0
13		**Contact**	Char	
14		**Format version**	Char	e.g., “HDF version 4.0”
15		**GHRSST format version**	Char	e.g. “GHRSST-PP data format v0,1”
16		**Conventions**	Char	"CF-1.0"

17		**Metadata_ID**	Char	ID of GHRSST-PP MDR record associated with this data file
18	**CSST**		short[*n*][*m*]	2-D array (*n* rows by *m* columns)
19		**long_name**	Char	*e.g.*"Collated sea surface temperature"
20		**Units**	Char	"Celsius"
21		**_FillValue**	Short	*e.g.* -32768
22		**add_offset**	Float	*e.g.* 0.0
23		**scale_factor**	Float	*e.g.* 0.01
24		**valid_min**	Same data type as variable	Minimum valid value for this variable
25		**valid_max**	Same data type as variable	Maximum value for this variable
26	**Grid_longitude**		Float[n]	Grid Cell longitude coordinate variable
27		**Units**	Char	"degrees_east"
28	**Grid_latitude**		Float[*m*]	Grid Cell latitude coordinante variable
29		**Units**	Char	"degrees_north"
30	**Confidence**		Double[*n*][*m*]	2-D array (*n* rows by *m* cols)
31		**Long_name**	Char	"Data processing, confidence and mask fags"
32		**Units**	Char	"none"
33		**Meaning**	Char	Bitfields –see data product specifications in sections 4-6
34		**_FillValue**	Byte	
35	**AdditionalConfidence**		Double[*n*][*m*]	2-D array (*n* rows by *m* cols)
36		**Long_name**	Char	"Data processing, confidence and mask fags"
37		**Units**	Char	"none"
38		**Meaning**	Char	Bitfields –see data product specifications in sections 4-6
39		**_FillValue**	Byte	

A1.2.1 GDSv1 netCDF global attributes

The global attributes shown in Table A1.6 are mandatory for each GDSv1 netCDF data product file. The optional CF convention attributes described in Table A1.7 are encouraged.

Table A1.6 Mandatory global attribute tags for GDSv1 netCDF data product files.

Tag name	Format	Description
`Title`	char[160]	A descriptive title for the data set
`entry_id`	char[200]	Filename suffix appended to Entry_ID for commensurate data set metadata entry, e.g. USA-RSS-AMSRE-MW-L2-SST.2004181.2315.amsr-e.sst.ncdf
`Location`	char [10]	GHRSST -PP data centre code (see Appendix 2 Table A2. 1)
`creation_date`	char[10],"yyyy-mm-dd"	Date the data file was created (UTC)
`start_time`	char[8],"hh:mm:ss" UTC	Start time of the data in universal time coordinated (UTC; ~ Greenwich Mean Time)
`stop_time`	char[8],"hh:mm:ss" UTC	End time of the data in universal time coordinated (UTC; ~ Greenwich Mean Time)
`start_date`	char[10],"yyyy-mm-dd" UTC	Start date of the datain universal time coordinated (UTC; ~ Greenwich Mean Time)
`stop_date`	char[10],"yyyy-mm-dd" UTC	End date of the data in universal time coordinated (UTC; ~ Greenwich Mean Time).

northernmost_latitude	Float	degrees north, range -90° to +90°
southernmost_latitude	Float	degrees north, range -90° to +90°
easternmost_longitude	Float	degrees east, range -180° to +180°
westernmost_longitude	Float	degrees east, range -180° to +180°
contact	Char	A free text string giving the primary contact for information about the data set
Format_version	Char	netCDF version 3.0
GDS_format_version	Char	v1.0
Conventions	Char	A text string identifying the netCDF convention followed. This attribute should be set to "CF-1.0" to indicate compatibility with the Climate and Forecast (CF) netCDF convention.
Metadata_ID	Char	ID of GHRSST-PP MDR record associated with this data file (from MMR)

Table A1.7 Optional global attribute tags for GDSv1 netCDF data product files.

Tag name	Format	Description
standard_name	Char	A standard and unique description of a physical quantity. For the complete list of standard name strings, see http://www.cgd.ucar.edu/cms/eaton/netcdf/standard_name.html.
Source	Char	Method of production of the original data
Institution	Char	Where the data was originally produced
References	Char	References that describe the data or the methods used to produce it. Include here the names of all GDS configuration files that have been used.
History	Char	List of the applications that have modified the original data
Axis	char[1]	Assigns a default axis direction to help applications decide how to plot the data. Possible values are X, Y, Z, or T.
Positive	Char	Indicates the real-world direction of a coordinate variable ("up" or "down"). *E.g.*, a depth or atmospheric pressure coordinate would have this attribute set to "down".
coordinates	Char	Identifies auxiliary coordinate variables, such as 2-D lat-lon coordinate arrays, label variables, and alternative coordinate variables.
Comment	Char	Miscellaneous information about the data or the methods used to produce it

A1.2.2 GDSv1 netCDF variable attribute definitions

A netCDF variable is a multidimensional array stored inside a netCDF file. The rank can be 1 (*i.e.* vector), 2 (*i.e.* array), or more as required. The variable name should be meaningful and concise. The optional `long_name` attribute is provided for a detailed and unambiguous description of the variable. Variables can be of any data type, and it is recommended to use the `add_offset` and `scale_factor` attributes to implement scaling functions to reduce the data volume. There may be many variables included within one GDSv1 data file as necessary (e.g., geolocation data, confidence flags and additional confidence flags as shown in Figure A1.1). It is recommended that array dimensions in C/C++ are ordered T, Z, Y, X, where X and Y are the two horizontal spatial dimensions, Z is the vertical, and T is time, with X the fastest varying dimension and T the slowest. FORTRAN arrays dimensions should be defined in the opposite order (X,Y,Z,T). Other dimensions should be added to the left of this list in C/C++, or to the right in FORTRAN.

The netCDF variable attributes described in Table A1.8 are mandatory and the optional netCDF variable tags described in Table A1.9 are encouraged.

Table A1.8 Mandatory variable attribute tags for GDSv1 netCDF data product files.

Tag name	Format	Description
`_FillValue`	Same data type as variable	A value used to indicate array elements containing no valid data
`Units`	char	Text description of the units, preferably S.I., and must be compatible with the Unidata UDUNITS package.

Table A1.9 Optional variable attribute tags for GDSv1 netCDF data product files.

Tag name	Format	Description
`Scale_factor`	Same data type as variable	To be multiplied by the variable to recover the original value
`add_offset`	Same data type as variable	To be added to the variable after multiplying by the scale factor to recover the original value. If only one of **scale_factor** or **add_offset** are needed, then both should be included anyway to avoid ambiguity, with **scale_factor** defaulting to 1.0 and **add_offset** defaulting to 0.0.
`long_name`	char	A long version of the variable name
`Meaning`	char	A free text description of the variable, providing more detailed information than the **long_name** attribute
`valid_min`	Same data type as variable	Minimum valid value for this variable
`valid_max`	Same data type as variable	Maximum valid value for this variable. The fill value should be outside this valid range

A1.2.3 GDSv1 coordinate Variable definitions

netCDF coordinate variables provide scales for the space and time axes for the multidimensional data arrays, and must be included for all dimensions that can be identified as spatio-temporal axes. Coordinate variables are of two types:

1. Coordinate vectors are used for data arrays located on orthogonal (but not necessarily regularly spaced) grids, such as a geographic (lat-lon) map projections. The only required attribute is **units**. The elements of a coordinate vector array should be in monotonically increasing or decreasing order. The data type can be any and scaling may be implemented if required.
2. Coordinate arrays are used to geolocate data arrays on non-orthogonal grids, such as images in the original pixel/scanline space, or complicated map projections. Required attributes are **units** and **_FillValue**. Elements of the coordinate array need not be monotonically ordered. The data type can be any and scaling may be implemented if required.

A1.3 GDSv1 WMO GRIB data format specification

Encoding some or all GHRSST products available in netCDF into GRIB format allows delivery through the Regional Meteorological Data Communication Network (RMDCN) and facilitates their use by meteorological services.

The World Meteorological Organization (WMO) Commission for Basic Systems (CBS) Extraordinary Meeting Number VIII (1985) approved a general purpose, bit-oriented data exchange format, designated FM 92-VIII Ext. GRIB (GRIdded Binary). It is an efficient vehicle for transmitting large volumes of gridded data to automated centers over high-speed telecommunication lines using modern protocols. GRIB can equally well serve as a data storage format, generating the same efficiencies relative to information storage and retrieval devices.

A complete description of the GRIB format can be found in WMO publication No 306, Manual on Codes. A guide to GRIB is available from WMO website at http://www.wmo.ch/web/www/WDM/Guides/Guide-binary-2.html

A GRIB encoding and decoding software is maintained and distributed by the European Center for Medium range Weather Forecast (ECMWF). All information about the software and its ordering is available from ECMWF website at http://www.ecmwf.int/products/data/software/grib.html

A.1.3.1 GRIB edition

Currently, GRIB format has two possible editions:

- edition 1, which is a WMO standard since 1991, and is widely used as format for NWP products distribution over RMDCN and through data (ftp) servers

- edition 2, which is also a WMO standard since November 2001, and which will not supersede edition 1 until several years

Edition 1 is limited to 16 MBytes for a single product. At grid description level, latitude-longitude grids are described with a 1/1000 degree accuracy. Decoding support is wide. Integration of GRIB edition 1 into viewers or processing is important for NMHSs and centres interfaced with meteorological/waves NWP data.

Edition 2 enables to handle up to 2**64 octets for a product that may be multi-field. Latitude-longitude grids are basically described with a 1/1000000 degree accuracy, but a fraction mesh is possible. Decoding support is still low. Though encoding to GRIB edition 2 should be considered, provision should be made to ensure the encoding from netCDF to GRIB edition 1. In the following, only the specifications for GRIB edition 1 are addressed.

A.1.3.2 GRIB structure and constraints

Each GRIB record intended for either transmission or storage contains a single parameter with values located at an array of grid points, or represented as a set of spectral coefficients, for a single level (or layer), encoded as a continuous bit stream. The logical divisions of the record are designated as "sections", each of which provides control information and/or data. A GRIB record consists of six sections, two of which are optional:

- (0) Indicator Section
- (1) Product Definition Section (PDS)
- (2) Grid Description Section (GDS) - optional
- (3) Bit Map Section (BMS) - optional
- (4) Binary Data Section (BDS)
- (5) '7777' (ASCII Characters)

Due to several features of the GRIB format, encoding GHRSST products to GRIB is not straightforward:

Octet 4 in Section 1 defines the Parameter Table Version number. Currently Version 3 is used for international exchange. Parameter table version numbers 128-254 are reserved for local use. To be widely useful a GRIB format should use Parameter Table version 3. For SST data the "temperature K" code can be used ; but there is no standard way through this table to distinguish between a collated, a merged or an analysed one. A workaround is to use the "Generating process ID number (allocated by the originating center)" in Section 1. Moreover no entry exists in Parameter Table Version 3 to identify auxiliary data types such as confidence flags or observation geolocation data. If they had to be encoded local codes would to be used.

A GRIB file can contain single-field data. Therefore related fields can't be linked in a formatted way.

Table A.1.3.1 GHRSST-PP products

Collated SST	Arguable	Generalization of GRIB format to all type of collated SST data is questionable (in particular to in situ measurements).
Merged SST	mandatory	
Analysed SST	mandatory	
Diurnal Variation	mandatory	
confidence flags	optional	not a standard parameter
observation geolocation data	optional	not a standard parameter

Encoding non standard parameters would need to choose carefully the "local" (128-254) for octet 9 of section 1. The following assumes that only the SST data fields are encoded.

A.1.3.3 Encoding a SST field

A.1.3.3.1 Representation of value

The representation of a single value is such that:

$$Y * 10^{D} = R + (X * 2^{E}) \qquad \textit{(Eqn 3.1.1.1)}$$

Y = original or unpacked value; units as in Table 2
D = decimal scale factor, to achieve desired precision(sign bit, followed by a 15-bit integer)
R = reference value (32 bits)
X = internal value (No. of bits varies for each record)
E = binary scale factor for variable bit word length packing (sign bit, followed by a 15-bit integer)
The reference value (R) is the minimum value of the (possibly) decimally scaled data that is being encoded. R is placed in the Binary Data Section in four octets as a single precision floating-point number :

sAAAAAAA BBBBBBBB BBBBBBBB BBBBBBBB

where

s = sign bit, encoded as : 0 means positive, 1 means negative
A...A = 7-bit binary integer, the characteristic
B...B = 24-bit binary integer, the mantissa

The appropriate formula to recover the value of R is:

$$R = (-1)^{s} * 2^{(-24)} * B * 16^{(A-64)} \qquad \textit{(Eqn. 3.1.1.2)}$$

Table A.1.3.2 Section 1

Octet no.	PDS Content	
1-3	Length in octets of the Product Definition Section	
4	Parameter Table Version number, currently 3 for international exchange. (Parameter table version numbers 128-254 are reserved for local use.)	3
5	Identification of center	TBD
6	Generating process ID number (allocated by the originating	TBD can be used to distinguish the level merged/analysed

	center)	
7	Grid Identification (geographical location and area, defined by the originating center)	255 (Grid is to be defined in Section 2)
8	Flag specifying the presence or absence of a GDS or a BMS	bits 1 and 2 = 1 (GDS and BMS included)
9	Indicator of parameter and units	11 (temperature: K)
10	Indicator of type of level or layer	102 (mean sea level)
11-12	Height, pressure, etc. of the level or layer	0
13	Year of century	Reference time UTC
14	Month of year	
15	Day of month	
16	Hour of day	
17	Minute of hour	
18	Time unit	1 (hour)
19	P1 - Period of time (Number of time units) (0 for analysis or initialized analysis.) Units of time given by content of octet 18	TBD can be used to code start date/time
20	P2 - Period of time (Number of time units) or Time interval between successive analyses, successive initialized analyses, or forecasts, undergoing averaging or accumulation. Units given by octet 18.	TBD can be used to code end date/time
21	Time range indicator	TBD
22-23	Number included in average, when octet 21 indicates an average or accumulation; otherwise set to zero.	TBD
24	Number Missing from averages or accumulations.	TBD
25	Century of Initial (Reference) time (=20 until Jan. 1, 2001)	
26	Identification of sub-center (allocated by the originating center)	TBD
27-28	The decimal scale factor D A negative value is indicated by setting the high order bit (bit No. 1) in octet 27 to 1 (on).	TBD
29-40	Reserved (need not be present)	not used
41-...	Reserved for originating center use.	not used

Table A.1.3.3: Section 2 (first part)

Octet no.	GDS Content	
1-3	Length in octets of the Grid Description Section	
4	NV, the number of vertical coordinate parameters	0
5	PV, the location (octet number) of the list of vertical coordinate parameters, if present or PL, the location (octet number) of the list of numbers of points in each row (when no vertical parameters are present), if present or 255 (all bits set to 1) if neither are present	255
6	Data representation type	0 Latitude/Longitude Grid - Equidistant Cylindrical or Plate Carree projection
7-32	Grid description, according to data representation	see below

Grid cell size is an open issue. Non decimal grid cell sizes raise potential problems. In GRIB edition 1, the only proper way to handle that would be not to document meshes ("increments"), with first "resolution flag" set to 0 (first bit of octet 17 within section 2, for a lat-long product). And preferably use decimal limits (exact in milli-degrees) for coordinates of first and last points. If the mesh was rounded and used in computations to have the position of a given point in the grid, the error would be significant. But GRIB (edition 1) lat-lon products with meshes unspecified are seldom used if ever. So there is a risk that already existing applications rely on meshes which must be explicitly specified.

Table A.1.3.4: Section 2: LATITUDE/LONGITUDE GRIDS INCLUDING GAUSSIAN

(GDS Octets 7 - 32)

OCTET NO.	CONTENT & MEANING	
7 - 8	Ni - No. of points along a latitude circle	
9 - 10	Nj - No. of points along a longitude meridian	
11 - 13	La1 - latitude of first grid point; units: millidegrees (degrees x 1000); values limited to range 0 - 90,000; bit 1 (leftmost) set to 1 for south latitude	
14 - 16	Lo1 - longitude of first grid point; units: millidegrees (degrees x 1000); values limited to range 0 - 360,000; bit 1 (leftmost) set to 1 for west longitude	

17	Resolution and component flags	bit 1 = 1 (Direction increments given) if Grid cell size is 1/10° bit 1 = 0 (Direction increments not given) if Grid cell size is 1/12° bit 2 = 1 Earth assumed oblate spheroid with size as determined by IAU in 1965: 6378.160 km, 6356.775 km, f = 1/297.0 bits 3-8 = 0
18 - 20	La2 - Latitude of last grid point (same units, value range, and bit 1 as La1)	
21 - 23	Lo2 - Longitude of last grid point (same units, value range, and bit 1 as Lo1)	
24 - 25	Di - Longitudinal Direction Increment (same units as Lo1) (if not given, all bits set = 1)	100 if grid cell size is 1/10° not given if 1/12°
26 - 27	Regular Lat/Lon Grid: Dj - Latitudinal Direction Increment (same units as La1) (if not given, all bits set = 1)	100 if grid cell size is 1/10° not given if 1/12°
28	Scanning mode flags	bit 1 = 0 Points scan in +i direction bit 2 = 0 Points scan in -j direction bit 3 = 0 Adjacent points in i direction are consecutive (FORTRAN: (I,J)) bits 4-8 =0 Note: i direction is defined as west to east along a parallel of latitude, or left to right along an x axis. j direction is defined as south to north along a meridian of longitude, or bottom to top along a y axis.
29 - 32	Reserved (set to zero)	

Table A.1.3.5: Section 3

Octet no.	**BMS Content**	
1-3	Length in octets of Bit Map Section	
4	Number of unused bits at end of Section 3.	
5-6	Numeric: 0 – a bit map follows otherwise the numerics refers to a prefenied bit map provided by the center	0
7-nnn	Bit map, zero filled to an even number of octets	where bits are 0 corresponds to netCDF _FillValue

Table A.1.3.6: Section 4

Octet no.	BMS Content	
1-3	Length in octets of binary data section	
4	Bits 1 through 4: Flag Bits 5 through 8: Number of unused bits at end of Section 4.	bit 1 = 0 Grid point data bit 2 = 0 Simple packing bit 3 = TBD (Original data were inter values / floating point values) bit 4 = 0 (No additional flags at octet 14)
5-6	The binary scale factor (E). A negative value is indicated by setting the high order bit (bit No. 1) in octet 5 to 1 (on).	TBD
7-10	Reference value (minimum value); floating point representation of the number.	TBD
11	Number of bits into which a datum point is packed	TBD
12 - nnn	Variable, depending on octet 4; zero filled to an even number of octets.	

A1.3.6 Concluding remarks

It is not possible to encode the GHRSST products into a GRIB edition 1 universal format because i) it is not possible to distinguish collated, merged and analyzed SSTs through standard description parameters, ii) confidence flags are not identified as standard parameters. They are ways to palliate this situation through using locally defined code values that represent deviations from the universal understanding of the format.

The example presented above of simply encoding the main SST fields has shown also that many options are left open and should be discussed by users.

As experienced by the EUMETSAT O&SI SAF the GRIB format is not well adapted to the variety of GHRSST products. It is thus recommended that encoding from netCDF to GRIB version 1, if considered as mandatory, be carefully examined by a panel of GRIB user's specialists before adoption.

A1.4 pre-processed (L2P) data products

L2P data products are derived from native format L2 data products. L2P data products consist of the original L2 data values that have been reformatted to a netCDF file format. A L2P confidence data record is provided for each L2 pixel, as described in A1.4.2.

A1.4.1 L2P filename convention.

The filename convention used for L2 pre processed SST (L2P) data products has been designed to provide useful information in an easily readable format. All L2P data product filenames are derived according to the following convention

<Processing Centre Code>--<L0 ID>->-<Date Valid>-<L2 filename>-<Processing Model ID>.<base format>

which is described in Table A1.4.1

Table A1.4.1 Collated data product filename components

Name	Definition	Description
<Processing Centre Code>	Defined in Appendix 2 Table A2.1	Processing centre code
<L0 ID>	Defined in Appendix 2 Table A2.2	Data set name
<Date Valid>	YYYYMMDD	Refers to the date for which this particular data set
<L2 filename>	Native to L2 filename	Filename of L2 data file as given by data provider
<processing model ID>	V1	refers to the GDS version number, in this case, V1
<base format>	Nc or grb	Generic file format (nc=netCDF, grb=WMO GRIB)

For example:

AUST-P-AVHRR16-L-20030621-LAC20030621A7SST-V1.0.nc

Refers to a data set having regional coverage that was generated at the Australian RDAC (AUST), it is a L2P pre-processed auxiliary data product (P), the source data is AVHRR NOAA 16 LAC (AVHRR16-L), it is valid on 21st June 2003 (20030621), has a native filename (LAC20030621A7SST) and was generated using GDS-v1 (V1.0).

A1.4.2 L2P pixel confidence data record format specification

Table A1.4.2 describes the L2P pixel confidence data record that should be created for each pixel.

Table A1.4.2 GDS L2P SST confidence data record format. Mandatory fields are shown in light blue [TBC at the GHRSST-PP Science Team workshop, September 2003]

Byte position	Name	Definition	Description	Units
+2bytes	**Time**	Julian days	Pixel acquisition time	UTC
+1byte	**ΔTmin**	WP-ID2.1.1.2	Deviation form reference SST	K, Scaled INT
+1byte	**SatZA**	WP-ID2.1.1,4	Solar zenith angle	Deg, Scaled INT
+1byte	**SolZA**	WP-ID2.1.1.5	Satellite Zenith angle	Deg, Scaled INT
+4bits	**SSI_value**	WP-ID2.1.1.15	Near contemporaneous SSI value	Wm^{-2}, Scaled into 150 Wm^{-2} intervals
+3bits	**Wspd**	WP-ID2.1.1.8	Wind speed value	Coded in 2ms-1 intervals: 0-2=0 2-4=1 4-6=2 6-8=3 8-10=4 >10=5 >15=6 no data=7
+2bit	**Aerosol_pres**	WP-ID2.1.1.12	Suspected	Code

	ent		atmospheric Aerosol present. 0=no aerosol, 1=native flag from L2 data stream, 3=reference data	
+1bit	**No_L2_data**	WP-ID2.1.1.1	No input data available	Flag
+1bit	**SST_Bad**	WP-ID2.1.1.2	SST out of limits	Flag
+4bits	**Land_sea_flag**	WP-ID2.1.1.11	Pixel type classification sea or land	Code
+1bit	**Cloud**	WP-ID2.1.1.13	Input data are all cloudy	Flag
+1bit	**SunGlint**	WP-ID2.1.1.6	Suspected sun glint: Glint > threshold	Flag
+1bits	**satZA_bad**	WP-ID2.1.1.4	satZA limit exceeded	Flag
+1bits	**solZA_bad**	WP-ID2.1.1.5	solZA limit exceeded	Flag
+1bit	**CosmeticFill**	WP-ID2.1.1.9	Pixel is a cosmetic value	Flag
+3bits	**FractionalSeaIce**	WP-ID2.1.1.10	Ice contamination flag 0=no sea ice, 1=sea ice flag from native L2 data stream, 3=sea ice from reference field	??
+2bits	**MW_SST_flag**	WP-ID2.1.1.7 and WP-ID2.1.1.8	MW SST tests failed (1=failed high u, 2=failed low u, 3=TMI failed low SST test)	Code, 0-3
+3bits	**Proximity_Confidence**	WP-ID2.1.1.14	Proximity confidence value	Code
+1bit	**L2_native_bias**	WP-ID2.1.1.15	Flag indicates bias error is provided by L2 native data	Flag
+1bit	**L2_native_rms**	WP-ID2.1.1.15	Flag indicates rms error is provided by L2 native data	Flag
+1byte	**bias**	WP-ID2.1.1.15	Error characteristics based on confidence flags	K, Scaled INT
+1bytes	**rms**	WP-ID2.1.1.15	Error characteristics based on confidence flags	K, Scaled INT
+3bits	**DV_confidence**	WP-ID3.2.4	Diurnal confidence	Code 0-7

			value	
+1bytes	**DV_bias**	WP-ID3.2.4	Estimated Diurnal magnitude	K, Scaled INT
+4bits	**DV_shape**	WP-ID3.2.5	Estimated shape of diurnal signal	Code
+2bytes	**Spare**		Reserved fro GDSv2	N/A

A1.5. Analysed L4 (L4) SST data products

L4 analysed data products, L4 are derived from an analysis procedure based on lower-level GDSv1 data products (depending on the analysis procedure that is used). L4 data products are formatted as netCDF and GRIB format data files and include confidence and error statistics for each grid-cell. L4 data products are produced for each APPW.

A1.5.1 L4 product filename convention

The filename convention used for L4 data products has been designed to provide useful information in an easily readable format. All L4 data product filenames are derived according to the following convention:

<Processing Centre Code>-<Product ID>-<Date Valid>-T<APPW>-<Processing Model ID>.<base format>

which is explained in Table A1.5.1

Table A1.5.1. L4 analysed data product filename components.

Name	Definition	Description
<Processing Centre Code>	Defined in Appendix 2 Table A2.1	Processing centre code
<Product ID>	Defined in Appendix 2 Table A2.2	Data set name
<Date Valid>	YYYYMMDD	Refers to the date for which this particular data set
<APPW>	00 or 12, Defined in Table 2.3.2	refers to the analysed product processing window (APPW) for which this particular data set is valid
<processing model ID>	V1	Refers to the GDS version number, in this case, V1
<base format>	nc or grb	Generic file format (nc=netCDF, grb=WMO GRIB)

For example:

JGDAC-FND-20030621T00-V1.0.grb

Refers to a data set having global coverage that was generated at the Jet Propulsion Laboratory GDAC (JGDAC), the data provides an estimate of the foundation SST (FND), it is valid on 21st June 2003 (20030621) and is valid for APPW 1 (T00). The product was generated using GDS-v1 (V1.0) and is formatted in GRIB data format (.grb).

A1.5.2. L4SST1m and L4SSTfnd grid cell confidence data record format specification

Table A1.5.2 describes format of GDSv1 L4 grid cell confidence values that should be created for each L4 grid cell.

Table A1.5.2 GDSv1 generic L4 SST product grid cell confidence data. Mandatory fields are shown with blue shading.

Byte position	Name	Definition	Description	Units
+2bit	**Aerosol_present**	WP-ID2.1.1.12	Suspected atmospheric Aerosol present in source data. 0=no aerosol, 1=native flag from L2 data stream, 3=reference data	Code
+1bit	**No_L2_data**	WP-ID2.1.1.1	No input data used in the analysis	Flag
+4bits	**Land_sea_flag**	WP-ID2.1.1.11	Grid cell type classification sea or land	Code
+??	**FractionalSeaIce**	WP-ID2.1.1.10	Ice contamination flag 0=no sea ice, 1=sea ice flag from native L2 data stream, 3=sea ice from reference field	Code
+2bits	**Grid_coverage**	WP-ID3.2.3	Estimated coverage of the grid cell area by input data	Code.
+3bits	**Wspd**	WP-ID2.1.1.8	Wind speed value	Coded in 2ms-1 intervals: 0-2=0 2-4=1 4-6=2 6-8=3 8-10=4 >10=5 >15=6 no data=7
+1byte	**Bias**	Rule 6.2.3	Error characteristics based on confidence flags and analysis scheme	K, Scaled INT
+1byte	**Rms**	Rule 6.2.3	Error characteristics based on confidence flags and analysis scheme	K, Scaled INT
+4bits	**SSI_value**	WP-ID2.1.1.15	Near contemporaneous SSI value	Wm^{-2}, Scaled into 150 Wm^{-2} intervals
+3bits	**Wspd**	WP-ID2.1.1.8	Wind speed value	Coded in 2ms-1 intervals: 0-2=0 2-4=1 4-6=2 6-8=3 8-10=4

				>10=5 >15=6 no data=7
+3bits	**DV_confidence**	WP-ID3.2.4	Diurnal confidence value	Code 0-7
+4bytes	**DV_bias**	WP-ID3.2.4	4 SSTskin estimates at 6 hourly intervals that can be used together with SST1m or SSTfnd to estimate diurnal variations throughout a day	K, Scaled INT
+4 bits	**DV_shape**	WP-ID3.2.5	Estimated general shape of diurnal signal	Code
+2bytes	**Spare**		Reserved fro GDSv2	N/A

Appendix 2. GDSv1 reference code tables

The following sections provide reference tables that are used within the GDSv1.

A2.1 GHRSST-PP data processing centre codes applicable to the GDSv1

The codes defined in Table A2.1 provide a unique identifier for each GHRSST-PP RDAC and GDAC data processing centre.

Table A2.1 Data centre prefix codes.

Prefix Code	Data centre name
MGDAC	US-GODAE Monterey GDAC
JGDAC	Jet Propulsion Laboratory GDAC
EUR	European RDAC
USA	United States RDAC
JAP	Japanese RDAC
SEASNET	SEASnet Tropical coverage RDAC
REMSS	Remote Sensing Systems, CA, USA
NASDA	National space Development Agency (of Japan)
ESA	European Space Agency
RSMAS	University of Miami, RSMAS.
TOHOKU	University of Tohoku, Japan
SOC	Southampton Oceanography Centre, UK
JPL	Jet Propulsion Laboratory
OSISAF	EUMETSAT Ocean and Sea Ice Satellite Applications Facility
AUST	Australian RDAC

A2.2 GDSv1 data set name codes

The codes defined in Table A2.2 provide a unique identifier for each data set referenced by the GDSv1.

Table A2.2 Data set name codes (see Appendix A3 for a full description)

L2 data stream	Description
ATS_NR__2P	ENVISAR AATSR near real time 1km SSTskin data
ATS_MET_2P	ENVISAT AATSR real time meteorological data product
AVHRR16_G	AVHRR NOAA-16 GAC derived SST data
AVHRR16_L	AVHRR NOAA-16 LAC derived SST data
AVHRR17_G	AVHRR NOAA-17 GAC derived SST data
AVHRR17_L	AVHRR NOAA-17 LAC derived SST data
SEVIRI	MSG-SEVIRI derived SST data
SEVIRI_SSI	MSG-SEVIRI derived SSI data
GOESE	GOES-E derived SST data
GOESE_SSI	GOES-E derived SSI data
GOESW	GOES-W derived SST data
GOESW_SSI	GOES-W derived SSI data
GMS	GMS derived SST data (GOES-9?)
GMS_SSI	GMS derived SSI data (GOES-9?)
AMSRE	AMSR-E derived SST data
AMSRE_WSP	AMSR-E derived wind speed data
AMSRE_WV	AMSR-E derived atmospheric water vapour data
AMSR	AMSR derived SST data
AMSR_WSP	AMSR derived wind speed data
AMSR_WV	AMSR derived atmospheric water vapour data

TMI	TRMM TMI derived SST data
TMI_WSP	TRMM TMI derived wind speed data
TMI_WV	TRMM TMI derived atmospheric water vapour data
TMI_VIRS	TRMM TMI-VIIRS derived SST data
AIRS	EOS-AQUA AIRS derived SST
MODIS_A	EOS AQUA MODIS derived SST data
MODIS_T	EOS TERRA MODIS derived SST data
AMSR_ICE	AMSR derived ocean- ice data product
AMSRE_ICE	AMSR-E derived ocean ice data product
SSMIFxx_WSP	SSMI derived wind speed (xx refers to the DMSP satellite number)
SSTFND	SSTfnd
SST1M	SST1m

Appendix 3. GDSv1P input data definitions

The following sub-sections collectively describe the data products that are used by the GDSv1.

A3.1 Reference data sets

The GDSv1 refers to the reference fields described in Table A3.1.1

Table A3.1.1.1Reference data fields used by the GDSv1

ID	Name	Description	Data agreement	Reference
R1	10 day Minimum Climatological SST1m	Ten year (1985-1995) time series of daily 9km Pathfinder SST data. The reference SST is derived as consecutive 10 day minimum night time SST maps		Faugere et al. (2001)
R2	GDS n-day mean SSTfnd	The mean SSTfnd computed for a n- day period. This product is computed from IDSI-PMv1 SSTfnd data products in real time each day	N/A	
R3	Maximum surface Solar irradiance	SSI: can be computed within the GDSv1 or can be a model output or can be computed from NWP fields (ECMWF)		NWP fields
R4	Daily mean wind speeds	Computed within GDSv1 merged auxiliary products or from NWP fields (ECMWF)		
R5	DMSP SSM/I Daily and Monthly Polar Gridded Sea Ice Concentrations.	NSIDC SSM/I sea ice products in polar stereographic projection currently include DMSP-F8, F11 and F13 daily and monthly sea ice concentrations. Data, gridded at a resolution of 25 x 25 km, begin 25 June 1987.		Cavialeri et al (1992) http://nsidc.org/data/nsidc-0002.html
R6	MODIS/Terra Land Cover Type 96-Day L3 Global 1km ISIN Grid	Land Cover Classification product, MOD12Q1, identifies 17 classes of land cover in the International Geosphere-Biosphere Programme (IGBP) global vegetation classification scheme. The data set is based on 1 year of MODIS data 2000-2001 and includes a land water cover QC flag that provides a simple ocean water classification.		http://edcdaac.usgs.gov/modis/mod12q1.htmll
R7	ISCCP Climatological Summary Products (D2) Monthly average of D1 quantities including mean diurnal cycle.	Climatological Summary Product - Revised algorithm (D2). Resolution: 280km equal-area grid, monthly, global coverage.		http://isccp.giss.nasa.gov/isccpframe.htmll

	Distribution and properties of total cloudiness and cloud types. Also includes snow/ice cover.			
R8	SOC Global Air-Sea Heat and Momentum Flux Climatology	Estimates of surface fluxes have been obtained from in situ reports within the COADS 1a (Comprehensive Ocean Atmosphere Dataset 1a), a global dataset containing of order 30 million surface observations from ships and buoys collected over the period 1980 - 93.		Josey et al. (1998) http://www.soc.soton.ac.uk/JRD/MET/fluxclimmon.php33
R9	Atmospheric Aerosol reference	NOAA product Current retrievals of aerosol optical depths (AOD) are made from AVHRR/3 reflectance channels 1 (0.63 um), 2 (0.83 um), and possibly 3A (1.61 um), from the two NOAA platforms: NOAA-16 (local equatorial crossing time, EXT~2 PM) and NOAA-17 (EXT~10 AM)		http://www.osdpd.noaa.gov/PSB/EPS/Aerosol/Aerosol.html
R10	OIv2: Reynolds Optimal Interpolated SST analysis v2	OI.v2 is a SST analysis produced by a blend of AVHRR and in situ data (ship and buoy). The analysis is produced operational weekly on a one-degree spatial grid.		Reynolds et al. (2002) http://www.emc.ncep.noaa.gov/research/cmb/sst_analysis/
R11	OI.v2 monthly climatology	monthly climatology produced from and other analyses with a base period of 1971-2000.		Xue et al (2003)) http://www.cpc.noaa.gov/products/predictions/30day/SSTs/sst_clim.html
R12	HadSST v1	Monthly 1 degree by 1 degree resolution SST and sea ice fields from HadISST1 from 1871 to date. Complete SST fields are reconstructed based on in situ measurements from ships and buoys through 1981 and from these blended with bias-corrected AVHRR SST from 1982 onwards		Rayner et al. (2003) http://www.metoffice.com/
R13	NWP data	Solar flux and wind speed required	ECMWF under GODAE agreement	

A3.2 L2 satellite SST data streams used in the GDSv1

The following sections provide a reference to each of the L2 data streams that will be used in the GDSv1. Table A3.2.1 provides a summary of each L2 data stream.

Table 3.2.1 Summary of L2 satellite SST sdata streams used by the GDSv1

Code	GHRSST-PP Identification Code	Sensor name	Nadir FoV resolution	Coverage (space and time)	Variable	Data provider	Data Format	GHRSST-PP ingestion	Data Transport	Applicable Data Agreement
1	**ATS_NR__2P**	AATSR	1km	Global with 3 day repeat	SSTskin	ESA	ATS_NR__2P (http://envisat.esa.int/dataproducts/aatsr/toc.htm)	EUR-RDAC GDAC	ftp pull from ESA/ESRIN	ESA Cat-1 proposal ID1329 (Donlon et al)
2	**ATS_MET_2P**	AATSR	10 arc min	Global, 3day repeat	SSTskin	ESA	ATS_MET_2P (http://envisat.esa.int/dataproducts/aatsr/toc.htm)	EUR-RDAC GDAC	ftp pull from ESA/ESRIN	ESA Cat-1 proposal ID1329 (Donlon et al)
3	**AVHRR16-G**	AVHRR NOAA 16	4 km (GAC)	Global, daily repeat	SST1m	NAVOCEANO		GDAC	ftp from US-GODAE	None
4	**AVHRR16-L**	AVHRR NOAA 16	1-2 km (LAC)	Regional, at least daily repeat, day and night	SST1m	1. NAVOCEANO		GDAC	ftp from US-GODAE	None
5	**AVHRR16-L**	AVHRR NOAA 16	2 km (LAC)	Regional, at least daily repeat, day and night	SSTsubskin	EUMESAT O&SI SAF	http://www.meteorologie.eu.org/safo	EUR RDAC	ftp from EUMESAT O&SI SAF ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI	None
6	**AVHRR17-G**	AVHRR NOAA 17	4 km (GAC)	Global, daily repeat	SST1m	NAVOCEANO		GDAC		NAVOCEANO
7	**AVHRR17-L**	AVHRR NOAA 17	1-2 km (LAC)	Regional, at least daily repeat, day and night	SST1mn	NAVOCEANO/		GDAC		NAVOCEANO
8	**AVHRR17-L**	AVHRR NOAA 17	1-2 km (LAC)	Regional, at least daily repeat, day and night	SSTsubskin	EUMESAT O&SI SAF	http://www.meteorologie.eu.org/safo	EUR RDAC	ftp from EUMESAT O&SI SAF ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI	None
9	**SEVIRI**	MSG	10 km	Regional, 3	SSTsubskin	EUMETSAT O&SI	http://www.meteorologie.e	EUR RDAC	ftp from EUMESAT O&SI	None

		SEVIRI		hour repeat		SAF	u.org/safo/gb_html_sst/regional/regsst_manual.pdf http://www.meteorologie.eu.org/safo/gb_html_sst/atlantic/atlsst_manual.pdf		SAF ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI	
10	**GOES-E**	GOES-8(12)	4 km	Regional, 3 hour repeat	SSTsubskin	EUMETSAT O&SI SAF	http://www.meteorologie.eu.org/safo/gb_html_sst/regional/regsst_manual.pdf http://www.meteorologie.eu.org/safo/gb_html_sst/atlantic/atlsst_manual.pdf	EUR RDAC	ftp from EUMESAT O&SI SAF ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI	None
11	**GOES-W**	GOES-10	4 km	Regional, ½ hour repeat	SSTsubskin ??					
12	**GOES-E**	GOES-8(12)	4 km	Regional, 1/2 hour repeat	SSTsubskin	NAVOCEANO				
13	**GOES-W**	GOES-10	4 km	Regional, ½ hour repeat	SSTsubskin ??	NAVOCEANO				
14	**GMS (GOES-9?)**	GMS	4 km	Regional, ½ hour repeat	SSTsubskin	U. Tohoku				
15	**GMS-SST (GOES-9?)**	GMS	4 km	Regional, ½ hour repeat	SST1m	U. Tohoku				
16	**AMSRE**	AMSR-E	25 km grid	Global, daily repeat	SSTsubskin	REMSS	http://www.remss.com/amsr/amsr_data_description.html	JPL GDAC	ftp from REMSS ftp.ssmi.com	Registration required
17	**AMSR**	AMSR	25 km grid	Global, daily repeat	SSTsubskin	NASDA				
18	**TMI**	TRMM-TMI	25 km grid	Global, daily repeat	SSTsubskin	REMSS	http://www.remss.com/tmi/tmi_browse.html	JPL GDAC	ftp from REMSS ftp.ssmi.com	None
19	**TMI-VIRS**	TRMM-VIRS	2km	Global, daily repeat	SSTsubskin	NASDA				NASDA/REMSS
20	**MODIS-A**	MODIS AQUA	4 km ?	Regional ??	SSTdepth					
21	**MODIS-T**	MODIS TERRA	4 km ?	Regional ??	SSTdepth				Real time service not currently available for the GDSv1 but expected in the GDSv2	

A3.2.1 Advanced Along Track Scanning Radiometer (AATSR) L2 GHRSST-PP data streams.

GDS	**ATS_NR__2P:** high resolution(1.1km) SSTskin data product **ATS_MET_2P:** low resolution (10 arc minute grid) averaged SSTskin and BT data product
Point of contact	Craig Donlon (craig.donlon@jrc.it, Tel: +39 0332 786353)
Data provider	European Space Agency (ESA)
Data transport	ftp pull service from GHRSST/ESA server
Data agreement	AATSR data are provided to the GHRSST-PP Science Team and other researchers free of charge within the framework of the GHRSST-PP under an ESA Category-1 data agreement ID is 1329 (see http://projects.esa-ao.org => ESA Data Policy=> Terms and conditions). Under this agreement, ESA make available in real time, to PIs named of Cat-1 1329. Additional researches can be added to this agreement by sending a signed copy of the ESA Category-1 data agreement and a short summary of the work that will be carried out using the AATSR data within the GHRSST-PP project.
Data products provided to GHRSST-PP	**ATS_NR__2P:** Gridded Surface Temperature (GST) Product at full resolution (1.1km). **ATS_MET_2P:** The Meteo product is a fast delivery product designed for use by meteorological offices, and contains averaged Brightness Temperatures and SST at 10 arc minute resolution. This is available in either ENVISAT format or BUFR format.
Data format document	A full specification of ENVISAT AATSR ATS_NR__2P and ATS_MET_2P data products can be found at http://envisat.esa.org/aatsr/. BUFR format http://envisat.esa.int/dataproducts/aatsr/CNTR6-1-2.htm#eph.aatsr.aatsrdf.2p.ATS_MET_2P
Primary GHRSST-PP entry point	EUR-RDAC (Medspiration Project), ESA ftp server for GHRSST-PP (METEO products only)
Secondary GHRSST-PP entry point	GDAC via ftp pull
Notes	Initially, only METEO data products will be available by ftp service.

A3.2.2 Advanced Very High Resolution Radiometer (AVHRR) L2 GHRSST-PP data streams.

GHRSST-PP ID	**AVHRR16-G**: GAC format NOAA 16 AVHRR data **AVHRR16-L**: LAC format NOAA 16 AVHRR data **AVHRR17-G**: GAC format NOAA 17 AVHRR data **AVHRR17-L**: LAC format NOAA 17 AVHRR data
Points of contact	1. Jorge Vasquez (jv@pacific.jpl.nasa.gov) Doug May (mayd@navocean.navy.mil) 2. Jorge Vasquez (jv@pacific.jpl.nasa.gov) Doug May (mayd@navocean.navy.mil) 3. Pierre LeBorgne (leborgne@meteo.fr) 4. Michel Petit (Michel.Petit@ird.fr)
Data provider	1. PO.DAAC and US-Navy 2. PO.DAAC and US-Navy 3. EUMETSAT O&SI SAF 4. Survey of Environment Assisted by Satellites (SEASnet) project
Data transport	1. L2-LAC 2.2 km MCSST (Regional coverage) ftp://podaac.jpl.nasa.gov/pub/sea_surface_temperature/avhrr/navoceano_hrpt_lac/data/L2 2. Orbital ungridded 9km MCSST

	ftp://podaac.jpl.nasa.gov/pub/sea_surface_temperature/avhrr/navoceano_mcsst/ 3. ftp pull from EUMETSAT O&SI SAF server ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI 4. ftp
Data agreement	1. None 2. None 3. None 4. TBD
Data products provided to GHRSST-PP	1. 2.2 km MCSST product using both Local Area Coverage (LAC) and High Resolution Picture Transmission (HRPT) SST calculations. Coverage includes the East and West Coasts of the United States, the Mediterranean, and parts of the Indian Ccean and the Australian Coasts. 2. Single orbit ungridded 9km Multichannel sea-surface temperatures (MCSST). Improved cloud retrieval algorithms applied by NAVOCEANO have significantly increased the number of SST pixels flagged as cloud free, as compared to previous MCSST data sets. 3. 3 hourly SSTsubskin (bias corrected using night time buoy data) stereo polar grid 4km resolution over the European Seas 4. SEASnet intend to produce GHRSST-PP L2P and L2C data products using LAC AVHRR SST in the tropical areas (Receiving stations: Reunion Island, French Guyana, Canary Islands, New Caledonia). Gridded data on a 0.01° angle. Coverage over LAC receiving stations (see ?? for image).
Data format document	1. http://podaac.jpl.nasa.gov/pub/sea_surface_temperature/avhrr/navoceano_mcsst/doc/avhrr_navoceano.html 2. http://podaac.jpl.nasa.gov/order/order_sstemp.html#Product143 3. Product manual at http://www.meteorlogie.eu.org/safo 4. netCDF (GHRSST-PP data format (see section 2.7)
Primary GHRSST-PP entry point	1. JPLGDAC 2. EUR RDAC 3. SEASnet RDAC 4. JAP RDAC 5. AUST RDAC
Secondary GHRSST-PP entry point	
Notes	1. The AVHRR sensor is used at all RDAC centres and contacts are given for each centre. 2. Additional LAC coverage data available at RDAC via local receiving stations

A3.2.3 METEOSAT Second Generation Spinning Enhanced Visible and Infrared Imager (SEVIRI) L2 GHRSST-PP data streams.

GHRSST-PP ID	**SEVIRI-SST**
Point of contact	Pierre LeBorgne (leborgne@meteo.fr)
Data provider	EUMETSAT O&SI SAF
Data transport	ftp pull from EUMETSAT O&SI SAF server ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI
Data agreement	None for EUMETSAT,
Data products provided to GHRSST-	Gridded SSTsubskin over the Atlantic Ocean from 100°W – 45°E and 60°N – 90°N at a grid resolution of 0.1°

PP	
Data format document	http://www.meteorologie.eu.org/safo/gb_html_sst/regional/regsst_manual.pdf http://www.meteorologie.eu.org/safo/gb_html_sst/atlantic/atlsst_manual.pdf
Primary GHRSST-PP entry point	EUR RDAC
Secondary GHRSST-PP entry point	
Notes	

A3.2.4 Geostationary Operational Environmental Satellite (GOES) 10 L2 GHRSST-PP data streams.

GHRSST-PP ID	**GOES-W**
Point of contact	Eileen Maturi (Eileen.mauturi@noaa.gov)
Data provider	NOAA/NESDIS
Data transport	ftp from JPL – currently in preparation
Data agreement	None
Data products provided to GHRSST-PP	[TBD]
Data format document	[TBD]
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	
Notes	

A3.2.5 Geostationary Operational Environmental Satellite (GOES) 8 (12) L2 GHRSST-PP data streams.

GHRSST-PP ID	**GOES-E-SST** **GOES-E-SSI**
Point of contact	1. Eileen Maturi (Eileen.mauturi@noaa.gov 2. Pierre LeBorgne (leborgne@meteo.fr)
Data provider	1. PO.DAAC at JPL 2. EUMETSAT O&SI SAF
Data transport	1. ftp pull from po.daac.jpl.nasa.gov 2. ftp pull from EUMETSAT O&SI SAF server ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI
Data agreement	1. None 2. None
Data products provided to GHRSST-PP	1. [TBD] 2. Gridded Surface Solar Irradiance (SSI) over the Atlantic Ocean from 100°W – 45°E and 60°N – 90°N at a grid resolution of 0.1° Data are available in real time at 3 hourly intervals.
Data format document	1. [TBD] 2. http://www.meteorologie.eu.org/safo/gb_html_ssi/atlantic/ssi_manual.pdf
Primary GHRSST-PP entry point	1. JPL GDAC 2. EUR RDAC
Secondary GHRSST-PP entry point	

Notes	

A3.2.6 Geostationary Meteorological Satellite (GMS) L2 GHRSST-PP data streams.

GHRSST-PP ID	**GMS-SST**
Point of contact	Hiroshi Kawamura (kamu@ ocean.caos.tohoku.ac.jp)
Data provider	University of Tohoku, Japan
Data transport	ftp pull
Data agreement	[TBD]
Data products provided to GHRSST-PP	[TBD]
Data format document	[TBD]
Primary GHRSST-PP entry point	Japanese RDAC
Secondary GHRSST-PP entry point	
Notes	1. GMS may be unavailable shortly

A3.2.7 Advanced Microwave Scanning Radiometer-E (AMSR-E) L2 GHRSST-PP data streams.

GHRSST-PP ID	**AMSRE: Gridded SST**
Point of contact	Chelle Gentemann (gentemann@remss.com)
Data provider	Remote Sensing systems (REMSS)
Data transport	ftp pull from ftp.ssmi.com/amsr
Data agreement	Agreement with GHRSST-PP Science Team for use of data products
Data products provided to GHRSST-PP	Each binary data file consists of twelve daily 0.25 x 0.25 degree grid (1440,720) byte maps. For daily files, six daytime maps in the following order, Time (UTC), Sea Surface Temperature (SST), 10 meter Surface Wind Speed (WSPD), Atmospheric Water Vapor (VAPOR), Cloud Liquid Water (CLOUD), and Precipitation Rate (RAIN), are followed by six nighttime maps in the same order.
Data format document	http://www.remss.com/amsr/amsr_data_description.html
Primary GHRSST-PP entry point	GDAC
Secondary GHRSST-PP entry point	EU RDAC others
Notes	

A3.2.8 Advanced Microwave Scanning Radiometer (AMSR) L2 GHRSST-PP data streams.

GHRSST-PP ID	**AMSR: Gridded SST**
Point of contact	Misako Katchi (ghrsst@eorc.nasda.go.jp)
Data provider	NASDA GHRSST-PP AMSR data server http://sharaku.eorc.nasda.go.jp/ADEOS2/ghrsst/index.html
Data transport	ftp helios.eorc.nasda.go.jp
Data agreement	Restricted to GHRSST-PP Science Team membership
Data products provided to GHRSST-PP	SST1m

Data format document	http://sharaku.eorc.nasda.go.jp/ADEOS2/ghrsst/index.html
Primary GHRSST-PP entry point	JAP RDAC
Secondary GHRSST-PP entry point	EUR RDAC
Notes	

A3.2.9 TRMM Microwave Imager (TMI) L2 GHRSST-PP data streams.

GHRSST-PP ID	TMI: Gridded SST
Point of contact	Chelle Gentemann (gentemann@remss.com)
Data provider	Remote Sensing Systems (REMSS)
Data transport	ftp pull from ftp.ssmi.com/tmi
Data agreement	None
Data products provided to GHRSST-PP	Daily binary data files consists of fourteen 0.25 x 0.25 degree grid (1440,320) byte maps. Seven ascending maps in the following order: Time (T), Sea Surface Temperature (SST), 10-meter Surface Wind Speed using 11 GHz (Z), 10-meter Surface Wind Speed using 37 GHz (W), Atmospheric Water Vapor (V), Cloud Liquid Water (L), and Precipitation Rate (R), are followed by seven descending maps in the same order.
Data format document	http://www.remss.com/tmi/tmi_description.html
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	EU RDAC, JAP RDAC
Notes	TMI data are also available from NSADA

3.2.10 TRMM Visible and Infrared Scanner (VIRS) L2 GHRSST-PP data streams.

GHRSST-PP ID	VIRS:
Point of contact	Hiroshi Kawamura (kamu@ ocean.caos.tohoku.ac.jp)
Data provider	NASDA currently in discussion
Data transport	[TBD]
Data agreement	[TBD]
Data products provided to GHRSST-PP	VIRS SST spanning ~40°N – 40°S spatial resolution of ~2 km at nadir. Contemporaneous with TMI data and extremely useful for understanding differences between IR and MW SST measurements
Data format document	[TBD]
Primary GHRSST-PP entry point	Japanese RDAC
Secondary GHRSST-PP entry point	Other RDAC/GDAC
Notes	TRMM VIRD data are currently in discussio

A3.2.11 MODIS L2 GHRSST-PP data streams.

GHRSST-PP ID	AQUA

	TERRA
Point of contact	Peter Minnett (pminnett@rsmas.miami.edu)
Data provider	[TBD]
Data transport	[TBD]
Data agreement	[TBD]
Data products provided to GHRSST-PP	Due to the large volume of data MODIS data products at high resolution are not foreseen to provide global coverage. Regional are coverage is possible. Furthermore, it is hoped that a reduced (5-10km) spatial resolution data product could be made available to the GHRSST-PP in real time in the future.
Data format document	[TBD]
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	Other RDAC
Notes	MODIS data are not currently available in real time and are thus not included in the GDSv1. However it is expected that MODIS data will be included in the GDSv2

A3.3 GDSv1 auxiliary satellite data streams

Table 3.3.1 provides a summary of the auxiliary satellite data streams that will be used in the GDSv1.

Table 3.3.1 Summary of input satellite auxiliary data streams considered by the GDS-v1.

Code	GHRSST-PP Identification Code	Sensor name	Nadir FoV resolution	Coverage (space and time)	Variable	Data provider	Data Format	GHRSST-PP ingestion	Data Transport	Applicable Data Agreement
101	NISE	SSM/I	25km	Global, daily	Sea Ice and extent	NSIDC	http://nsidc.org/data/docs/daac/nise1_nise.gd.html	JPL GDAC	ftp pull	Registration required
102	AMSRE-ICE	AMSR-E	25 km	Daily	Sea Ice extent	REMSS	http://www.remss.com/amsr	JPL GDAC	ftp pull ftp.ssmi.com	Registration required (brewer@remss.com)
103	SEVIRI-SSI	MSG SEVIRI	0.1 x 0.1 degree	3 hourly	SSI	EUMETSAT O&SI SAF	http://www.meteorologie.eu.org/safo/gb_html_ssi/atlantic/ssi_manual.pdf	EUR RDAC	ftp pull from EUMETSAT O&SI SAF server ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI	None
104	GOESW-SSI									
105	GOESE-SSI									
106	GMS-SSI	GMS								
107	AMSRE-WSP	AMSR-E	0.25° grid	Gloabl	Gridded surface (10m) wind speed	REMSS	http://www.remss.com/amsr/amsr_data_description.html	JPL GDAC	ftp pull ftp.ssmi.com	Registration required (brewer@remss.com)
108	AMSRE-WV	AMSR-E	0.25° grid	Global	Gridded atmospheric water vapour	REMSS	http://www.remss.com/amsr/amsr_data_description.html	JPL GDAC	ftp pull ftp.ssmi.com	Registration required (brewer@remss.com)
109	AMSRE-CLD	AMSR-E	0.25° grid	Global	Gridded cloud liquid water vapour	REMSS	http://www.remss.com/amsr/amsr_data_description.html	JPL GDAC	ftp pull ftp.ssmi.com	Registration required (brewer@remss.com)

110	AMSR-WSP	AMSR		Gloabl	Gridded surface (10m) wind speed					
111	AMSR-WV	AMSR		Global	Gridded atmospheric water vapour					
112	AMSR-CLD	AMSR		Global	Gridded cloud liquid water vapour					
113	TMI-WSP	TRMM TMI	0.25° grid	40°N – 40°S, 180E - 180W	Gridded surface (10m) wind speed	REMSS	http://www.remss.com/tmi/tmi_data_description.html	JPL GDAC	ftp pull ftp.ssmi.com/tmi	none
114	TMI-WV	TRMM TMI	0.25° grid	40°N – 40°S, 180E - 180W	Gridded atmospheric water vapour	REMSS	http://www.remss.com/tmi/tmi_data_description.html	JPL GDAC	ftp pull ftp.ssmi.com/tmi	none
115	TMI-CLD	TRMM TMI	0.25° grid	40°N – 40°S, 180E - 180W	Cloud liquid water vapour	REMSS	http://www.remss.com/tmi/tmi_data_description.html	JPL GDAC	ftp pull ftp.ssmi.com/tmi	None
116	SSMI-WSP	DMSP SSM/I	0.25° grid	Global	Griddde surface (10m) wind speed	REMMS	http://www.remss.com/ssmi/ssm_data_description.html	JPL GDAC	ftp pull ftp.ssmi.com/ssmi	None

A3.3.1 National Snow and Ice Data Centre SSMI/I sea Ice products L2 GHRSST-PP data streams.

GHRSST-PP ID	**NISE**
Point of contact	Donald Cavalieri (Donald.J.Cavalieri@nasa.gov)
Data provider	National Snow and ice data Center (NSIDC, http://nsidc.org)
Data transport	ftp pull but not yet real time
Data agreement	Registration required at NSIDC
Data products provided to GHRSST-PP	Near Real-Time SSM/I EASE-Grid Daily Global Ice Concentration and Snow Extent product (Near real-time Ice and Snow Extent, NISE) provides daily, global near real-time maps of sea ice concentrations and snow extent. The National Snow and Ice Data Center (NSIDC) created the NISE product using passive microwave data from the Defense Meteorological Satellite Program (DMSP) F13 Special Sensor Microwave/Imager (SSM/I). Sea ice concentration and snow extent maps are provided in two 25 km azimuthal, equal-area projections: the Southern Hemisphere 25 km low resolution (Sl) and Northern Hemisphere 25 km low resolution (Nl) Equal-Area Scalable Earth-Grids (EASE-Grids). Data in Hierarchical Data Format - Earth Observing System (HDF-EOS) format, and browse files in GIF and HDF formats, are updated daily and are available via ftp for two weeks after initial posting.
Data format document	http://nsidc.org/data/docs/daac/nise1_nise.gd.html
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	
Notes	This is a secondary data stream as the AMSR-E 12.5 km sea ice data products are preferred. However these are not finalised or available as of Feb 2003

A3.3.2 AMSR-E Sea Ice L2 GHRSST-PP data streams.

GHRSST-PP ID	**AMSRE-ICE**
Point of contact	M Brewer (brewer@remss.com)
Data provider	Remote Sensing Systems
Data transport	ftp pull from ftp.ssmi.com
Data agreement	Registration required
Data products provided to GHRSST-PP	0.25° x 0.25° data as part of AMSRE SST bmaps_v2.0 format data. Provides ice extent.
Data format document	REMSS bmaps_v2 format. http://www.remss.com/amsr/amsr_data_description.html
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	RDAC by direct ftp pull
Notes	It is expected that this data stream will be superseded by 12.5 km AMSR and AMSR-E sea ice data products for the GDSv2

A3.3.3 METEOSAT Second Generation Spinning Enhanced Visible and Infrared Imager (SEVIRI) L2 GHRSST-PP data streams.

GHRSST-PP ID	**SEVIRI-SSI**
Point of contact	Pierre LeBorgne (leborgne@meteo.fr)
Data provider	EUMETSAT O&SI SAF
Data transport	ftp pull from EUMETSAT O&SI SAF server ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI
Data agreement	None for EUMETSAT,
Data products provided to GHRSST-PP	Gridded Surface Solar Irradiance (SSI) over the Atlantic Ocean from 100°W – 45°E and 60°N – 90°N at a grid resolution of 0.1°
Data format document	Product manual at http://www.meteorlogie.eu.org/safo
Primary GHRSST-PP entry point	EUR RDAC
Secondary GHRSST-PP entry point	US-GODAE via IFREMER ftp push
Notes	

A3.3.4 Geostationary Operational Environmental Satellite (GOES) 10 L2 GHRSST-PP data streams.

GHRSST-PP ID	**GOES-W-SSI**
Point of contact	Eileen Maturi (Eileen.mauturi@noaa.gov)
Data provider	NOAA/NESDIS
Data transport	ftp from JPL – currently in preparation
Data agreement	None
Data products provided to GHRSST-PP	[TBD]
Data format document	[TBD]
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	)
Notes	

A3.3.5 Geostationary Operational Environmental Satellite (GOES) 8(12) L2 GHRSST-PP data streams.

GHRSST-PP ID	**GOES-E-SSI**
Point of contact	1. Eileen Maturi (Eileen.mauturi@noaa.gov) 2. Pierre LeBorgne (leborgne@meteo.fr)
Data provider	1. NOAA/NESDIS 2. EUMETSAT O&SI SAF
Data transport	1. ftp pull from jpl GDAC 2. ftp pull from EUMETSAT O&SI SAF server ftp://ftp.ifremer.fr/pub/ifremer/cersat/SAFOSI
Data agreement	1. [TBD] 2. None
Data products	1. [TBD] 2. Gridded Solar Irradiance (SSI) over the Atlantic Ocean from 100°W – 45°E

provided to GHRSST-PP	and 60°N – 90°N at a grid resolution of 0.1°
Data format document	1. [TBD] 2. http://www.meteorologie.eu.org/safo/gb_html_ssi/atlantic/ssi_manual.pdf
Primary GHRSST-PP entry point	1. JPL GDAC 2. EUR RDAC
Secondary GHRSST-PP entry point	
Notes	

A3.3.6 Geostationary Meteorological Satellite (GMS) L2 GHRSST-PP data streams.

GHRSST-PP ID	**GMS-SSI**
Point of contact	Hiroshi Kawamura (kamu@ ocean.caos.tohoku.ac.jp)
Data provider	University of Tohoku, Japan
Data transport	ftp pull
Data agreement	[TBD]
Data products provided to GHRSST-PP	[TBD]
Data format document	[TBD]
Primary GHRSST-PP entry point	Japanese RDAC
Secondary GHRSST-PP entry point	
Notes	1. GMS may be unavailable shortly

A3.3.7 Advanced Microwave Scanning Radiometer-E (AMSR-E) L2 GHRSST-PP data streams.

GHRSST-PP ID	**AMSRE-WSP: Gridded surface (10m) wind speed** **AMSRE-WV: Gridded atmospheric water vapour** **AMSRE-CLD: Gridded cloud liquid water vapour**
Point of contact	Chelle Gentemann (gentemann@remss.com)
Data provider	Remote Sensing systems (REMSS)
Data transport	ftp pull from ftp.ssmi.com/amsr
Data agreement	Agreement with GHRSST-PP Science Team for use of data products
Data products provided to GHRSST-PP	Each binary data file consists of twelve daily 0.25 x 0.25 degree grid (1440,720) byte maps. For daily files, six daytime maps in the following order, Time (UTC), 10 meter Surface Wind Speed (WSPD), Atmospheric Water Vapor (VAPOR), Cloud Liquid Water (CLOUD), and Precipitation Rate (RAIN), are followed by six nighttime maps in the same order.
Data format document	http://www.remss.com/amsr/amsr_data_description.html
Primary GHRSST-PP entry point	GDAC
Secondary GHRSST-PP entry point	EU RDAC others
Notes	

A3.3.8 Advanced Microwave Scanning Radiometer (AMSR) L2 GHRSST-PP data streams.

GHRSST-PP ID	**AMSR-WSP: Gridded surface (10m) wind speed** **AMSR-WV: Gridded atmospheric water vapour** **AMSR-ICE: Sea ice concentration**
Point of contact	Misako Kachi (ghrsstpp@eorc.nasda.go.jp)
Data provider	NASDA
Data transport	[TBD]
Data agreement	[TBD]
Data products provided to GHRSST-PP	[TBD]
Data format document	[TBD]
Primary GHRSST-PP entry point	JAP RDAC
Secondary GHRSST-PP entry point	Other RDAC
Notes	Currently in discussion

A3.3.9 TRMM Microwave Imager (TMI) L2 GHRSST-PP data streams.

GHRSST-PP ID	**TMI-WSP: Gridded surface (10m) wind speed** **TMI-WV: Gridded atmospheric water vapour** **TMI-CLD: Cloud liquid water vapour**
Point of contact	Chelle Gentemann (gentemann@remss.com)
Data provider	Remote Sensing Systems (REMSS)
Data transport	ftp pull from ftp.ssmi.com/tmi
Data agreement	TBD
Data products provided to GHRSST-PP	Daily binary data files consists of fourteen 0.25 x 0.25 degree grid (1440,320) byte maps. Seven ascending maps in the following order: Time (T), 10-meter Surface Wind Speed using 11 GHz (Z), 10-meter Surface Wind Speed using 37 GHz (W), Atmospheric Water Vapor (V), Cloud Liquid Water (L), and Precipitation Rate (R), are followed by seven descending maps in the same order.
Data format document	http://www.remss.com/tmi/tmi_description.html
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	EU RDAC, JAP RDAC
Notes	1. TMI data only cover 40°N – 40°S in low earth orbit configuration 2. TMI data are also available from NSADA

A3.3.10 SSM/I surface (10m) wind speed L2 GHRSST-PP data streams.

GHRSST-PP ID	**SSMI-WSP**
Point of contact	D. Smith (smith@remss.com)
Data provider	Remote Sensing Systems (REMSS)
Data transport	ftp pull from ftp.ssmi.com/tmi
Data agreement	None
Data products	Daily binary data files consists of fourteen 0.25 x 0.25 degree grid 0.25 x

provided to GHRSST-PP	0.25 degree grid (1440,720) byte maps. For daily files, five morning maps in the following order, Time (T), 10 meter Surface Wind Speed (W), Atmospheric Water Vapor (V), Cloud Liquid Water (L), and Precipitation Rate (R)
Data format document	http://www.remss.com/ssmi/ssmi_description.html
Primary GHRSST-PP entry point	JPL GDAC
Secondary GHRSST-PP entry point	EU RDAC, JAP RDAC
Notes	

A3.4 GDSv1 in situ data streams

Table 3.4.1 provides a summary of the in situ data that are available to the GHRSST-PP in near real time (2-3 day delay).

Table 3.4.1 Summary of input in situ data streams considered by the GDS-v1.

GHRSST-PP Identification Code	Platform	Variables	Data provider	Data Format	GHRSST-PP ingestion	Data Transport	Applicable Data Agreement
TAO	Tropical Pacific Moored buoy array (TAO)	SST1m, SSTdepth, + others$	PMEL	http://www.pmel.noaa.gov/tao/index.shtml	GAC and RDAC	GTS	
PIRATA	Tropical Atlantic Moored buoy	SST1m, SSTdepth, + others$	PMEL/AOML/IFREMER	http://www.pmel.noaa.gov/pirata/	GAC and RDAC	GTS	
ARGO	Profiling float	SSTdepth	CORIOLIS data centre (IFREMER)	http://www.ifremer.fr/coriolis/cdc/default.htm		GDS and ftp	None
RADIOM	Radiometer	SSTskin	None specific	Various	RDAC	ftp	Permission
CORIOLIS-SHIP CORILIS-PROFILE	Ships, profiles, buoys	Various	CORIOLIS data centre (IFREMER)	http://www.ifremer.fr/coriolis/cdc/default.htm	EUR RDAC	GTS and ftp	None
MEDS-DRIFT MEDS-MOOR MEDS-PROFILE	Drifting buoy	SST1m	Fisheries and Oceans service, Canada	http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Home_e.htm	GDAC/RDAC	GTS and ftp	None
JMA	Ships, profiles, buoys	SST1m	Japanese Meteorological Agency	TBC	RDAC	TBC	TBC

$ wind speed, air temperature, atmospheric surface pressure, short-wave radiation

A3.4.1 Tropical Atmosphere Ocean (TAO) buoys L2 GHRSST-PP data streams.

GHRSST-PP ID	**TAO**
Point of contact	atlasrt@noaa.gov
Data provider	Pacific Marine Environmental Research Laboratory (PMEL)
Data transport	GTS see http://www.wmo.ch/web/www/TEM/gts.html or by ftp from PMEL see http://www.pmel.noaa.gov/tao/proj_over/availability.html for availability
Data agreement	None
Data products provided to GHRSST-PP	Wind speed, wind direction, Air temperature, relative humidity, rainfall, Short wave radiation, long wave radiation, air pressure, SST at various depths, salinity see http://www.pmel.noaa.gov/tao/proj_over/sensors.shtml for full details of sensors and available data.
Data format document	ASCII or netCDF data files available on demand
Primary GHRSST-PP entry point	TBD
Secondary GHRSST-PP entry point	
Notes	

A3.4.2 Pilot Research Moored Array in the Tropical Atlantic (PIRATA) buoys L2 GHRSST-PP data streams.

GHRSST-PP ID	**PIRATA**
Point of contact	Jaques Servain (jaques.servain@ird.fr) http://www.brest.ird.fr/pirata/piratafr.html
Data provider	Pacific Marine Environmental Research Laboratory (PMEL)
Data transport	GTS see http://www.wmo.ch/web/www/TEM/gts.html or by ftp from PMEL see http://www.pmel.noaa.gov/tao/proj_over/availability.html for availability http://www.pmel.noaa.gov/tao/data_deliv/deliv-pir.html
Data agreement	None
Data products provided to GHRSST-PP	Wind speed, wind direction, Air temperature, relative humidity, rainfall, http://www.pmel.noaa.gov/pirata/pir_statgus.html for full details of sensors and available data.
Data format document	http://www.pmel.noaa.gov/tao/data_deliv/deliv-pir.html
Primary GHRSST-PP entry point	TBD
Secondary GHRSST-PP entry point	
Notes	

A3.4.3 ARGO profiling floats L2 GHRSST-PP data streams.

GHRSST-PP ID	**ARGO**
Point of contact	Sylvie Pouliquen
Data provider	1. **CORIOLIS** data centre, IFREMER, Brest France. (http://www.coriolis.eu.org/coriolis/cdc/) 2. **US-GODAE** server (http://www,usgodae.org/argo/argo.html)
Data transport	ftp pull

Data agreement	TBD
Data products provided to GHRSST-PP	**Vertical profiles of SSTdepth (below 5m), pressure and conductivity** Argo is a global array of 3,000 free-drifting profiling floats that will measure the temperature and salinity of the upper 2000 m of the ocean. This will allow continuous monitoring of the climate state of the ocean, with all data being relayed and made publicly available within hours after collection. Argo deployments began in the year 2000. The Argo array is part of the Global Climate Observing System/Global Ocean Observing System (GCOS/GOOS) and part of the Climate Variability and Predictability Experiment (CLIVAR) and the Global Ocean Data Assimilation Experiment (GODAE). GHRSST-PP will use ARGO data within the top 10m of the sea surface contributing to the MDB and validation systems
Data format document	Argo data users manual (available at http://www.coriolis.eu.org/cdc/argo_rfc.htm)
Primary GHRSST-PP entry point	RDAC/GDAC
Secondary GHRSST-PP entry point	
Notes	

A3.4.4 Radiometer L2 GHRSST-PP data streams.

GHRSST-PP ID	**RADIOM:**
Point of contact	Peter Minnett, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway,Miami, FL 33149-1098, USA (email: pminnett@rsmas.miami.edu, Tel: Tel: +1 (305) 361-4104)
Data provider	Peter Minnett (pminnett@rsmas.miami.edu) • M-AERI measurements Craig Donlon (craig.donlon@jrc.it) • ISAR radiometer data in E-channel/Bay of Biscay Ian Robinson (ian.s.robinson@soc.soton.ac.uk) • ISAR radiometer data in E-channel/Bay of Biscay Andrew Jessup (jessup@apl.washington.edu) • CIRIMS radiometer data Ian Barton (ian.barton@marine.csiro.au) • DAR-011 and TASCO measurements Tim Nightingale (tim.nightingale@rl.ac.uk) • SISTeR measurements
Data transport	ftp pull by arrangement with individual PIs.
Data agreement	Several informal and formal agreements may be required between GHRSST-PP and the individual PI's responsible for the collection and processing of in situ radiometer data sets.
Data products provided to GHRSST-PP	SSTskin and other parameters
Data format document	Various data formats associated with different radiometer data sets. Ideally the following GHRSST-PP radiometer ASCII csv data format is proposed: yyyy,mm,dd,hh,mm,ss,latitude,longitude,SSTskin,SSTdepth,wind_speed, solar_rdadiation[,optional other data]

Primary GHRSST-PP entry point	RDAC
Secondary GHRSST-PP entry point	
Notes	Unprocessed radiometer data may be difficult to use. A real time data service for certain deployments in place (Caribbean, soon in Bay of Biscay/E. Channel) and other NRT services are in preparation (trans-Atlantic)

A3.4.5 CORIOLIS GHRSST-PP data streams.

GHRSST-PP ID	**CORIOLIS-SHIP** **CORIOLIS-MOORINGS**
Point of contact	Sylvie Pouliquen
Data provider	CORIOLIS data centre, IFREMER, Brest France. (http://www.coriolis.eu.org/coriolis/cdc/)
Data transport	ftp pull /GTS ??
Data agreement	TBD
Data products provided to GHRSST-PP	
Data format document	
Primary GHRSST-PP entry point	
Secondary GHRSST-PP entry point	
Notes	

A3.4.6 MEDS L2 GHRSST-PP data streams.

GHRSST-PP ID	**MEDS-DRIFTER** **MEDS-SHIP** **MEDS-PROFILE**
Point of contact	Estelle Couture (couture@meds-sdmm.dfo-mpo.gc.ca, Tel: +613 990-0259
Data provider	Marine Environmental Data Service, Department of Fisheries and Oceans Canada, 12W082 - 200 Kent Street, Ottawa, Ontario, K1A 0E6, Canada.
Data transport	GTS
Data agreement	TBD
Data products provided to GHRSST-PP	**SST1m, sea level pressure, surface current velocity** Global sea surface meteorological and oceanographic observations reported in daily and historical time frames. Drifting buoy data includes the buoy position, date, time and in most cases includes many variables such as surface and subsurface water temperature, air pressure, air temperature, wind speed and wind direction.
Data format document	MEDS is the world data centre for drifting buoys (Responsible National Oceanographic Data Centre - RNODC). As part of its role, MEDS acquires, processes, quality controls and archives real-time drifting buoy messages reporting over the Global Telecommunications System (GTS) as well as delayed mode data acquired from other sources. The real time data available from MEDS that include SST measurements

	are: SST: -Drifting buoys http://www.meds-sdmm.dfo-mpo.gc.ca/alphapro/rnodc/main_glob_e.shtml -Thermosalinographs on ships of opportunity http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Databases/TrackOB/Trackob_e.htm Temperature profiles: -BATHYs and TESACs http://www.meds-sdmm.dfo-mpo.gc.ca/meds/Databases/OCEAN/Realtime_e.htm Data will be delivered in 3 times per week to an FTP site (TBD) and in the format of your choice, either ASCII or Net CDF.
Primary GHRSST-PP entry point	TBD
Secondary GHRSST-PP entry point	TBD
Notes	Some duplication may occur between CORIOLIS and MEDS

Appendix 4 GDSv1 Validation match-up database (MDB)

A match-up database (MDR) of near contemporaneous in situ, L2 SST and GDSv1 SST data products will be generated. The MDB will also include GDSv1 auxiliary data sets and additional in situ observations. The MDB will be used to validate GDSv1 data products and to derive appropriate quantitative values for rms. and bias estimates that can be attributed to GDS grid cell confidence codes. Each RDAC and GDCA centre will be responsible for preparing GDSv1 MDB records as described in WP-ID8

This section presents the format of the GHRSST-PP validation records. There are two types of validation data records depending on whether the validation is being done against an in situ measurement or against a spatial field, such as an SST field derived from a previously-validated satellite product – the latter applies to the DDS system.

A4.1 MDB validation records for in situ comparisons

A list of the variables that should comprise each validation record is provided in Table A4.1.1. This is based on experience gained with the AVHRR and MODIS validation exercises.

Table A4.1.1 Definition of in situ comparison MDB validation records.

Rcode	Record type, based on source of satellite data, and type of SST retrieval, or merged or analysed product, and type of validation measurement.
DDSID	ID code for Diagnostic Data Set (0 for not a DDS)
Insitudate	Date (yymmdd) of the in situ measurement
Insitutime	Time (hhmmss) of the in situ measurement
Satdate	Continuous time coordinate (seconds from January 1, 1981)
Lat	Satellite latitude
Lon	Satellite longitude
Solz	Solar zenith angle
Satz	Satellite zenith angle
Azim	Satellite azimuth angle
Glintf	Sun glint flag
Qual	Quality flags for brightness temperatures of pixels
Qsst	Quality flag for SST of pixel
Nbt	Number of brightness temperature channels used in the SST derivation, and included in this record. If none available, then N=0.
C	Pixel values (brightness temperatures, if available, and SST retrieval n 5x5 box)
Med	Median value of 5x5 box
Min	Minimum of 5x5 box
Max	Maximum of 5x5 box
Avg	Average of 5x5 box
Sd	Standard deviation of 5x5 box
Bdate	Continuous time coordinate of reference measurement (seconds from January 1, 1981)
Blat	Reference measurement latitude
Blon	Reference measurement longitude
Bid	Reference measurement id
Bsst	Reference measurement temperature
Oisst or SSTreference	Background SST for central pixel (e.g. Reynolds OI) or use GHRSST-PP reference SST [TBC]
Wind	Wind speed measured at central pixel

Windf	Wind speed flag giving provenance of data, and quality
Pw	Precipitable water measured at central pixel
Pwf	Precipitable water flag giving provenance of data, and quality
Aeros	Aerosol index measured at central pixel
Aerosf	Aerosol index flag giving provenance of data, and quality
Other information	e.g., MODIS Mirror angle of incidence (if appropriate) and Mirror side used (If appropriate)

A4.2 MDB validation records for satellite-to-satellite comparisons

A list of the variables that should comprise each validation record is provided in Table A4.2.1, in which the 'reference measurement' refers to the reference satellite-SST field, and the 'satellite measurement' means the satellite-SST field being validated. The area may cover a DDS, a specific area not confined to a DDS, or the entire global ocean. The fields must be represented in the standard GHRSST projection so there is a simple direct mapping from each element in one array to the same element in another array. The time difference between the measurements of the satellite and reference fields is an important parameter in the validation. For fields derived from single sensors on single satellites it is possible to generate the time interval between the measurements given the vectors of the sub satellite tracks as a function of time. For sensors on the same satellite, this is of course trivial. For fields derived by merging data from several satellites there may be no unique time attributable to individual cells, in which case a nominal average time for the whole data set is acceptable.

Table A4.2.1 Definition of MDB-FR satellite-to-satellite validation records.

Rcode	Record type, based on sources of satellite data, and type of SST retrieval, or merged or analysed product
DDSID	ID code for Diagnostic Data Set (0 for not a DDS; 999 for global)
Refstartdate	Date (yymmdd) of the first reference measurement
Refstarttime	Time (hhmmss) of the first reference measurement
Refenddate	Date (yymmdd) of the last reference measurement
Refendtime	Time (hhmmss) of the last reference measurement
Refdirn	Flag to show whether reference data from ascending or descending arcs, or from a combination
Refeqtime	Nominal equator crossing time of the ascending node for the reference data
Refsubsat	Vector of sub-satellite track as time, latitude longitude at intervals not to exceed one minute.
Refconvert	Expression for converting stored values to reference SST (°C)
Measstartdate	Date (yymmdd) of the first satellite measurement
Measstarttime	Time (hhmmss) of the first satellite measurement
Measenddate	Date (yymmdd) of the last satellite measurement
Measendtime	Time (hhmmss) of the last satellite measurement
Measdirn	Flag to show whether satellite data from ascending or descending arcs, or from a combination
Measeqtime	Nominal equator crossing time of the ascending node for the satellite data
Meassubsat	Vector of sub-satellite track of measurements as time, latitude longitude at intervals not to exceed one minute.
Measconvert	Expression for converting stored values to satellite SST (°C)
Nlines	Number of lines in image data array
Npix	Number of pixels in image data array
NWlat	Latitude of centre of pixel in NW corner of array

NWlon	Longitude of centre of pixel in NW corner of array
Dx	X increment (E-W) between pixels in image data array (km)
Dy	Y increment (N-S) between lines in image data array (km)
RefSST	Array of reference SST (Nlines lines x Npix pixels)
Refflag	Array of quality flag values for reference SST (Nlines lines x Npix pixels)
MeasSST	Array of measured SST (Nlines lines x Npix pixels)
Measflag	Array of quality flag values for measured SST (Nlines lines x Npix pixels)
Naux	**Number of auxiliary data arrays (such as winds, precipitable water….)**
Auxname_i	**Name and units of i^{th} auxiliary data array**
Auxconvert_i	**Expression for converting stored values to i^{th} auxiliary units**
Auxval_i	**Values of i^{th} auxiliary data array (Nlines lines x Npix pixels)**
Auxflag_i	**Quality flag for i^{th} auxiliary data array (Nlines lines x Npix pixels)**
	<last 4 records repeated Naux times>

A4.3 XML document Type Definition for GHRSST-PP MDB records

All GHRSST-PP MDB data records are encoded using the extensible mark-up language (XML). XML provides an application independent way of sharing and controlling the quality and content of data. A detailed description of XML-v1.0 can be found at http://www.w3.org/TR/REC-xml

Figure A4.3.1 XML DSD for GHRSST-PP MDB records [TBD at the GHRSST-PP Workshop, September 2003]

```
<?xml version="1.0"?>
<!—DTD for GHRSST-PP MDB records v1.0 26th May 2003-->
<!—C J Donlon -->
<!DOCTYPE MDB [
     TBD after final MMR data format is agreed at GHRSST-PP
     Science Team Workshop, Spetember 2003>
]
```

Notes:

- <!ELEMENT element-name (#PCDATA)> means that the element contains data that is going to be parsed by the MMR parser.
- <!ELEMENT element-name (#CDATA)> means that the element contains data that is not going to be parsed by the MMR parser.

Appendix 5. GHRSST-PP High Resolution Diagnostic Data Set (HR-DDS) data granule data product format specification

The GHRSST-PP High-resolution Diagnostic Data Set (HR-DDS) system provides a distributed data resource and a framework to analyse and inter-compare L2P and L4 data products in near real time, together with other data products including NWP analyses, operational ocean and atmospheric model outputs and in situ observations. In order to reduce the overhead of data storage and transport, the HR-DDS consists of about150 globally distributed 2° x 2° latitude x longitude sites. Figure A5.1 shows a map of primary[10] HR-DDS sites (v2.3) for which all L2P and L4* data streams[11] will be extracted and archived as ***data granules***. Each HR-DDS site is strategically positioned in order to address a particular issue; for example, a particularly dynamic or "quiet" oceanographic or atmospheric region, location of additional in situ infrastructure, areas known to be influenced by atmospheric aerosol or persistent cloud cover and areas already having a significant scientific interest. Each site shown in Figure A5.1 is defined in Section A5.3.

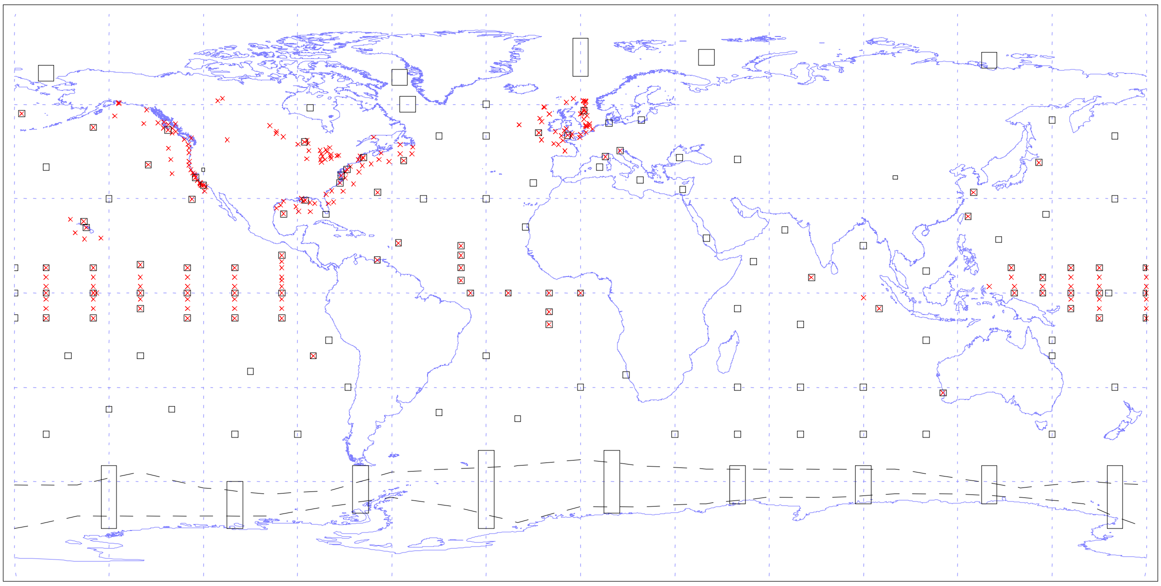


Figure A5.1 Location of HR-DDS sites v2.3 (April 2003) Based on output of the 2nd & 3rd GHRSST-PP workshop Science Team feedback. Moorings and permanent in situ installations are indicated by red asterisk (note that some are located on inland lakes). The HR-DDS is fully documented in the HR-DDS Implementation Plan (GHRSST/14).

The HR-DDS system provides a real time data resource the main learning tool for GHRSST-PP project scientists to develop and refine the GDS, to investigate differences between complementary data streams, to investigate regional and time variant bias statistics, to monitor the quality of input and output data streams and as a core component of the GHRSST-PP reanalysis (RAN) project. The HR-DDS constitutes the virtual laboratory of the GHRSST-PP allowing, for example, a full implementation of

[10] In addition, other temporary DDS sets can be established that correspond to cruise tracks or other areas of interest (for example local scale projects in specific areas)

[11] Note that Native L2 data streams will also be extracted for certain sensors and HR-DDS locations but these are not discussed in the GDS which focuses on the use of L2P and L4 GDSv1 data products only.

Integrated Global Observing Strategy (IGOS) measurement principles (see http://www.igospartners.org/).

This Appendix provides a summary overview of the GHRSST-PP HR-DDS system. A complete Scientific and technical description of the HR-DDS system can be found in GHRSST/13 available at http://www.ghgrsst-pp.org.

A5.1 Summary of the HR-DDS system configuration and operation

HR-DDS granules are produced by RDAC and GDAC centres in real time as netCDF data files according to the specifications provided in Appendix 1. HR-DDS granules are archived locally together with a corresponding HR-DDS metadata record. HR-DDS metadata records are produced according to the specifications described in Appendix A6 and are identical to other GDS metadata records. HR-DDS archive centres are referred to as HR-DDS nodes and in general, HR-DDS nodes should be part of RDAC and GDAC.

Figure A5.2 provides a schematic diagram of the GHRSST-PP HR-DDS data system which is based on a number of distributed HR-DDS nodes that are interconnected using the Distributed Oceanographic Data System (DODS, see http://www.dods.org).

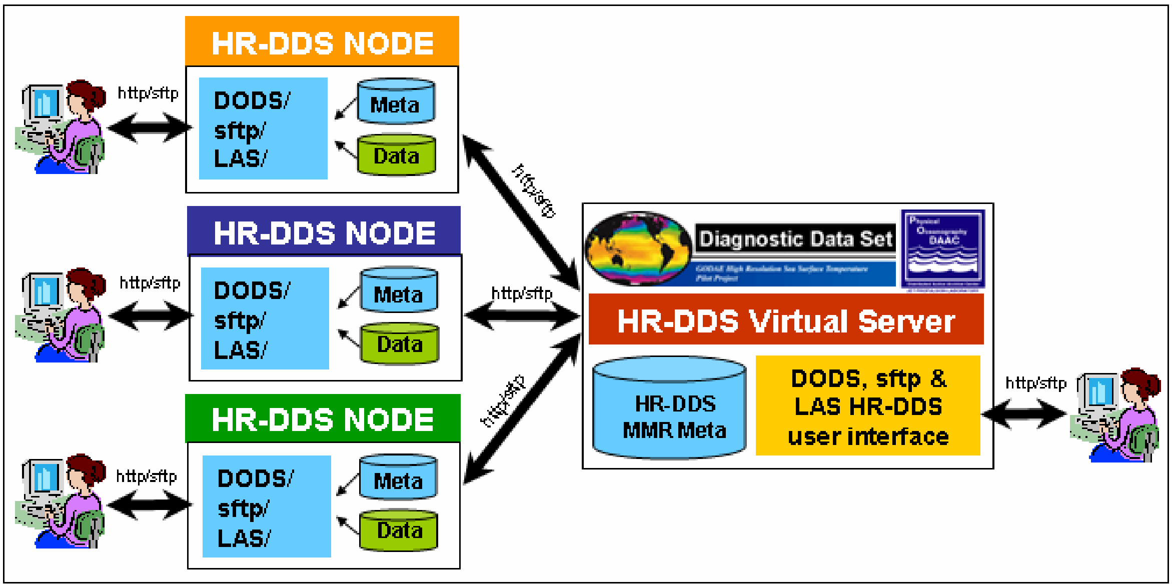


Figure A5.2 Schematic diagram of the GHRSST-PP HR-DDS system. Users may access HR-DDS data through local HR-DDS nodes or the global distributed data archive via a HR-DDS virtual server that coordinates GHRSST-PP HR-DDS data through the HR-DDS MMR.

The DODS architecture uses a client/server model and the http protocol to provide a framework that simplifies all aspects of scientific data networking. It provides tools (such as DODS servers) that make local data accessible to remote locations regardless of local storage format. HR-DDS granules at HR-DDS nodes may be accessed directly using a number of DODS aware client applications i(e.g., IDL, MATLAB) from any location. The Open source project to develop and extend a data access protocol (OPeNDAP) has now evolved from the DODS.

In addition to local data access via DODS/OPeNDAP servers at each HR-DDS node, HR-DDS nodal data archives are virtually combined as a single data archive by a HR-DDS virtual server. The HR-DDS virtual server uses GHRSST-PP Master Metadata

Repository (MMR) HR-DDS metadata records to catalogue HR-DDS data granules archived at ***all*** distributed HR-DDS nodes. Using a HR-DDS interface, users at the HR-DDS virtual server may access all HR-DDS granules (maintained at distributed locations, served by local DODS servers) using DODS aware client applications as if they were a single data archive residing on a local computer.

In addition to DODS access, HR-DDS granules may also be accessed using secure ftp (sftp) transactions or through a Live Access Server (LAS, http://ferret.pmel.noaa.gov/Ferret/LAS/). LAS is a highly configurable Web server designed to provide flexible access to geo-referenced scientific data and can present distributed data sets as a unified virtual data base through the use of DODS networking Using a web browser interface LAS enables a user to

- visualize data with on-the-fly graphics
- request custom subsets of variables in a choice of file formats
- access background reference material about the data (metadata)
- compare (difference) variables from distributed locations
- LAS enables the data provider to
- unify access to multiple types of data in a single interface
- create thematic data servers from distributed data sources
- offer derived products on the fly
- remedy metadata inadequacies (poorly self-describing data)
- offer unique products (e.g. visualization styles specialized for the data)

LAS will be used extensively in the GODAE project and will allow data to be extracted from many operational ocean model systems for inclusion and comparison with HR-DDS SST data. In addition, the LAS system is in active development to provide advanced features such as temporal aggregation of data and advanced display options in collaboration with the GODAE Data Sharing Pilot Project.

Each HR-DDS node is thus responsible for:

(a) Installing and maintaining a data archive of HR-DDS granules and associated MMR DSR and MMR-FS metadata records (see Appendix 6)
(b) Operating a DODS server to serve HR-DDS data
(c) Operating an sftp server for HR-DDS granule access
(d) Operating other optional data interface software (e.g., LAS)

A5.2 HR-DDS data granule file format

HR-DDS data granules should be extracted for all L2P and L4* data products in real time. Each HR-DDS data product should be formatted as a netCDF data file following the formatting description given for each GDSv1 data family in Appendix 1.

A5.3 HR-DDS metadata record format and registration

HR-DDS granule metadata records are identical in content and format to GDSv1 metadata records. A full description of MMR-DSR and MMR-FR is given in Appendix 6.

A MMR DSR will be registered at the GHRSST-PP MMR for each HR-DDS site that is described in Appendix A5.4. A MMR-FR will be generated for each HR-DDS granule and registered at the MMR following the procedure described in Appendix A6.4.

A5.4 Location of HR-DDS sites and responsible RDAC

Tables A5.4.1, A5.4.2, A5.4.3 and A5.4.4 describe the locations and formal names of each HR-DDS (v2) site for open ocean, moored buoys, SIMBIOS and other sites respectively.

Table A5.4.1 GODAE/GHRSST-PP Diagnostic Data Set (DDS) Open Ocean coordinates v2.3 (April 2003) Based on output of the 2nd & 3rd GHRSST-PP workshop Science Team feedback.

Location (GCMD location_valid)	GHRSST short name)	Latitude (lower left)	Longitude (lower left)	Δy-size	Δx-size	Description	Comment
GHRSST	ghr001	35	-15	2.0	2.0	Atlantic Ocean Madeira	
GHRSST	ghr002	50	-30	2.0	2.0	N Atlantic	
GHRSST	ghr003	30	-30	2.0	2.0	Central N Sub-tropical Atlantic	
GHRSST	ghr004	-20	-30	2.0	2.0	Central S Sub-tropical Atlantic	
GHRSST	ghr005	-62.5	-30	25.0	5.0	Weddel Sea (ice edge)	
GHRSST	ghr006	50	-45	2.0	2.0	W N Atlantic	
GHRSST	ghr007	30	-50	2.0	2.0	W Atlantic	
GHRSST	ghr008	-38	-45	2.0	2.0	Argentina coast	
GHRSST	ghr009	76.5	-74	2.0	2.0	North Water Polynya	
GHRSST	ghr010	-62.5	-70	15.0	5.0	Drake passage (ice edge)	
GHRSST	ghr011	-30	-74	2.0	2.0	Chile	
GHRSST	ghr012	-45	-90	2.0	2.0	Pacific-Antarctic basin	
GHRSST	ghr013	-25	-105	2.0	2.0	Easter Island	
GHRSST	ghr014	-45	-110	2.0	2.0	Pacific Antarctic Ridge	
GHRSST	ghr015	-67.5	-110	15.0	5.0	Amudsen Sea (ice edge)	
GHRSST	ghr016	-37	-130	2.0	2.0	Pitcairn Island S Pacific	
GHRSST	ghr017	30	-150	2.0	2.0	Central N Pacific	
GHRSST	ghr018	-37	-150	2.0	2.0	S Central Pacific	
GHRSST	ghr019	-65	-150	20.0	5.0	W Ross Sea (ice edge)	
GHRSST	ghr020	70	-170	5.0	5.0	Chukchi Sea	
GHRSST	ghr021	40	-170	2.0	2.0	Alutian Islands	
GHRSST	ghr022	-45	-170	2.0	2.0	Kermadec trench	S Pacific
GHRSST	ghr023	50	170	2.0	2.0	W Alutian Islands	
GHRSST	ghr024	30	170	2.0	2.0	NW Pacific	
GHRSST	ghr025	-30	170	2.0	2.0	Norfolk Is. S Pacific	
GHRSST	ghr026	-65	170	20.0	5.0	E Ross Sea (ice edge)	

GHRSST	ghr027	55	150	2.0	2.0	Sea of Okhotsk	
GHRSST	ghr028	-15	150	2.0	2.0	Coral Sea	
GHRSST	ghr029	-45	150	2.0	2.0	S Tasman Ridge	
GHRSST	ghr030	-61	130	12.0	5.0	S Ocean (ice edge)	
GHRSST	ghr031	7	110	2.0	2.0	S China Sea	
GHRSST	ghr032	-15	110	2.0	2.0	Java Trench	
GHRSST	ghr033	-45	110	2.0	2.0	Eastern S Ocean	
GHRSST	ghr034	15	90	2.0	2.0	Bay of Bengal	
GHRSST	ghr035	-30	90	2.0	2.0	Central S. Ocean	
GHRSST	ghr036	-45	90	2.0	2.0	Central S. Ocean	
GHRSST	ghr037	-61	90	12.0	5.0	S Ocean (ice edge)	
GHRSST	ghr038	20	65	2.0	2.0	Arabian Sea	
GHRSST	ghr039	10	55	2.0	2.0	Somali Jet	
GHRSST	ghr040	-10	70	2.0	2.0	Indian Ocean	
GHRSST	ghr041	-30	70	2.0	2.0	S Indian Ocean	
GHRSST	ghr042	-45	70	2.0	2.0	Kerguelen Is. S Ocean	
GHRSST	ghr043	-5	50	2.0	2.0	Somali Basin	
GHRSST	ghr044	-30	50	2.0	2.0	Mauritius Basin	
GHRSST	ghr045	-45	50	2.0	2.0	S Indian Ocean	
GHRSST	ghr046	-61	50	12.0	5.0	S Indian Ocean (ice edge)	
GHRSST	ghr047	-45	30	2.0	2.0	Agulhas Basin	
GHRSST	ghr048	36	19	2.0	2.0	Mediterranean	
GHRSST	ghr049	-30	0	2.0	2.0	SE Atlantic	
GHRSST	ghr050	-60	10	20.0	5.0	SE Atlantic (ice edge)	
GHRSST	ghr051	60.0	-55.0	5.0	5.0	Davis Strait	
GHRSST	ghr052	60.0	-30.0	2.0	2.0	Denmark Strait	
GHRSST	ghr053	43.0	31.5	2.0	2.0	Black Sea	
GHRSST	ghr054	42.5	50.0	2.0	2.0	Caspian Sea	
GHRSST	ghr055	39.2	-120.0	1.0	1.0	Lake Tahoe	USA
GHRSST	ghr056	36.8	100.2	1.0	1.5	Qinghai Hu Lake	China
GHRSST	ghr057	17.5	40.0	2.0	2.0	Red Sea	
GHRSST	ghr058	-15.0	-80.0	2.0	2.0	Peru current	
GHRSST	ghr059	-20.0	-85.0	2.0	2.0	IMET/EPIC mooring under stratus cloud	
GHRSST	ghr060	40.0	6.0	2.0	2.0	Western Mediterranean	
GHRSST	ghr061	-20.0	150.0	2.0	2.0	Great Barrier Reef	

GHRSST	ghr062	59.0	-86.0	2.0	2.0	Hudson Bay	
GHRSST	ghr063	75.0	0.0	12.0	5.0	Greenland Sea and ice edge	
GHRSST	ghr064	25.0	-81.0	2.0	2.0	Florida Keys (C-MAN sites)	
GHRSST	ghr065	-40.0	-20.0	2.0	2.0	S Atlantic	
GHRSST	ghr066	75.0	40.0	5.0	5.0	Barents Sea	
GHRSST	ghr067	74.0	130.0	5.0	5.0	Laptev Sea	
GHRSST	ghr068	-20.0	-140.0	2.0	2.0	South Pacific	

Table A5.4.2 GODAE/GHRSST-PP Diagnostic Data Set (DDS) Moored buoy site coordinates v2.3 (April 2003) Based on output of the 2nd & 3rd GHRSST-PP workshop Science Team feedback.

Location (GCMD location_valid)	GHRSST short name)	Latitude (lower left)	Longitude (lower left)	Δy-size	Δx-size	Description	Comment
NDBC-46035	ndb051	57.08	-177.71	2.0	2.0	Bering Sea	
TAO	tao038	8.0	-170.0	2.0	2.0		(Aug 1992)
TAO	tao044	-8.0	-170.0	2.0	2.0		(Aug 1992)
TAO	tao041	0.0	-170.0	2.0	2.0		(May 1998)
NDBC-46066	ndb063	52.65	-155.00	2.0	2.0	Kodiak	
TAO	tao031	8.0	-155.0	2.0	2.0		(Aug 1992)
TAO	tao037	-8.0	-155.0	2.0	2.0		(Mar 1992)
NDBC-46006	ndb039	40.84	-137.49	2.0	2.0	SW. Astoria	
TAO	tao018	8.0	-125.0	2.0	2.0		(Oct 1992)
TAO	tao021	0.0	-125.0	2.0	2.0		(Oct 1983)
TAO	tao024	-8.0	-125.0	2.0	2.0		(Sep 1992)
TAO	tao025	9.0	-140.0	2.0	2.0		(May 1988)
TAO	tao028	0.0	-140.0	2.0	2.0		(Apr 1983)
TAO	tao030	-5.0	-140.0	2.0	2.0		(Oct 1990)
TAO	tao011	8.0	-110.0	2.0	2.0		(Oct 1991)
TAO	tao014	0.0	-110.0	2.0	2.0		(Jan 1979)
TAO	tao017	-8.0	-110.0	2.0	2.0		(Nov 1985)
TAO	tao001	12.0	-95.0	2.0	2.0		(Dec 1999)
TAO	tao007	0.0	-95.0	2.0	2.0		(Jul 1981 inactive 83- 92)
TAO	tao010	-8.0	-95.0	2.0	2.0		(Aug 1994)
CAN-46147	can001	51.83	-131.22	2.0	2.0	Vancouver	

MF-41100	mfr001	15.9	-57.9	2.0	2.0	Antilles (Meteo France)	
CAN-44141	can002	42.10	-56.22	2.0	2.0	Gulf Stream	
EGOS-62081	ego003	51.0	-13.3	2.0	2.0	Celtic Sea	
PIRATA-Reggae	pir007	15.0	-38.0	2.0	2.0		(Jan 1998)
PIRATA-Forro	pir008	12.0	-38.0	2.0	2.0		(Feb 1999)
PIRATA-Lambada	pir009	8.0	-38.0	2.0	2.0		(Jan 1998)
PIRATA-Frevo	pir010	4.0	-38.0	2.0	2.0		(Feb 1999)
PIRATA-Samba	pir006	0.0	-35.0	2.0	2.0		(Jan 1998)
PIRATA-Jazz	pir005	0.0	-23.0	2.0	2.0		(Mar 1999)
PIRATA-Soul	pir001	0.0	0.0	2.0	2.0		(Feb 1998)
PIRATA-Java	pir002	0.0	-10.0	2.0	2.0		(Sep 1997)
PIRATA-Valse	pir003	-6.0	-10.0	2.0	2.0		(Mar 2000)
PIRATA-Gavotte	pir004	-10.0	-10.0	2.0	2.0		(Sep 1997)
UKMO-63117	ukm004	58.0	1.10	2.0	2.0	North Sea	
TAO/TRITON	tao073	-5.0	95.0	2.0	2.0		(26 Oct 2001)
NDBC-45001	ndb027	48.06	-87.78	2.0	2.0	Mid Lake Superior	USA
TAO/TRITON	tao069	8.0	137.0	2.0	2.0		(28 Sep 2001)
TAO/TRITON	tao071	0.0	138.0	2.0	2.0		(03 Oct 2001)
TAO/TRITON	tao065	5.0	147.0	2.0	2.0		(Feb 1990; Triton from 1999)
TAO/TRITON	tao067	0.0	147.0	2.0	2.0		(Apr 1994; Triton from 1999)
TAO/TRITON	tao059	8.0	156.0	2.0	2.0		(Dec 1994; Triton from 1999)
TAO/TRITON	tao062	0.0	156.0	2.0	2.0		(Jul 1995; Triton from 1999)
TAO/TRITON	tao064	-5.0	156.0	2.0	2.0		(Aug 1991; Triton from 1999)
TAO	tao055	0.0	165.0	2.0	2.0		(Jan 1986)
TAO	tao052	8.0	165.0	2.0	2.0		(Jul 1989)
TAO	tao058	-8.0	165.0	2.0	2.0		(Aug 1991)
TAO	tao045	8.0	180.0	2.0	2.0		(Nov 1993)
TAO	tao048	0.0	180.0	2.0	2.0		(Mar 1993)
TAO	tao051	-8.0	180.0	2.0	2.0		(Nov 1993)
NDBC-42002	ndb008	25.17	-94.42	2.0	2.0	Western Gulf	

Table A5.4.3 GODAE/GHRSST-PP Diagnostic Data Set (DDS) SIMBIOS Diagnostic Data Set Site coordinates http://simbios.gsfc.nasa.gov/Info/STM2001/Sites.html .

Location (GCMD location_valid)	GHRSST short name)	Latitude (lower left)	Longitude (lower left)	Δy-size	Δx-size	Description	Comment
SIMBIOS-MOBY	sim001	20.8	-157.20	2.0	2.0	Hawaii	
SIMBIOS-BATS/BTM	sim002	32.0	-64.50	2.0	2.0	Bermuda	
SIMBIOS-CALCOFI	sim003	29.85	-123.59	2.0	2.0	California	
SIMBIOS-EqPAC	sim004	0.0	-155.00	2.0	2.0	Eastern Equ Pacific	
SIMBIOS-HOT	sim005	22.75	-158.00	2.0	2.0	Hawaii	
SIMBIOS-Ishigaki	sim006	24.39	123.27	2.0	2.0	East China Sea	
SIMBIOS-Ligurian_Sea	sim007	43.37	7.90	2.0	2.0	Mediterranean	
SIMBIOS-Lower_Chesapeake_Bay	sim008	37.40	-76.13	2.0	2.0	Virginia	
SIMBIOS-Monterey_Bay	sim009	36.75	-122.42	2.0	2.0	Monterey	
SIMBIOS-Plymbody	sim010	50.2	-4.10	2.0	2.0	English Channel	
SIMBIOS-Venice_Tower	sim011	45.31	12.60	2.0	2.0	Northern Adriatic	
SIMBIOS-Station_H	sim012	41.5	145.78	2.0	2.0	Japan East Coast	
SIMBIOS-Cariaco_Basin	sim013	10.5	-64.66	2.0	2.0	Venezuela	
SIMBIOS-Kashidoo	sim014	4.95	73.45	2.0	2.0	Maldives Islands	
SIMBIOS-Korean	sim015	32.0	125.00	2.0	2.0	East China Sea	
SIMBIOS-LEO_15	sim016	39.3	-74.25	2.0	2.0	New Jersey	
SIMBIOS-Plumes_and_Blumes	sim017	34.25	-119.92	2.0	2.0	off Santa Barbara CA	
SIMBIOS-Scotian_Prince_Route	sim018	43.00	-69.00	2.0	2.0	Gulf of Maine	
SIMBIOS-NOAA-GOM	sim019	29.50	-87.50	2.0	2.0	Northern Gulf of Mexico	
SIMBIOS-NOAA-NC	sim020	35.00	-76.50	2.0	2.0	off North Carolina	
SIMBIOS-Rottnest_Island	sim021	-31.80	115.30	2.0	2.0	off Western Australia	

Table A5.4.4 GODAE/GHRSST-PP Diagnostic Data Set (DDS) general site coordinates v2.3 (April 2003) Based on output of the 2nd & 3rd GHRSST-PP workshop Science Team feedback.

Location (GCMD location_valid)	GHRSST short name)	Latitude (lower left)	Longitude (lower left)	Δy-size	Δx-size	Description	Comment
SIMBIOS-Arm_1	sim022	0.00	168.00	2.0	2.0	Nauru Island	
SIMBIOS-Arm_2	sim023	25.00	148.00	2.0	2.0	Manus Island	

SIMBIOS-NW_Afr_Upwell	sim025	21.00	-17.50	2.0	2.0	Morocco	
SIMBIOS-Alberon_Gyre	sim026	33.00	32.50	2.0	2.0	Eastern Mediterranean	
SIMBIOS-Helgoland	sim027	54.00	9.00	2.0	2.0	North Sea	
SIMBIOS-Nordic	sim028	55.00	19.30	2.0	2.0	Baltic Sea	
SIMBIOS-Luderitz_Upwell	sim029	-26.00	14.50	2.0	2.0	Namibia	
SIMBIOS-Philippine_Sea	sim030	17.00	133.00	2.0	2.0	Southeast Asia	
SIMBIOS-Cook_Islands	sim031	-20.00	-163.00	2.0	2.0	Australia	

Appendix 6. GHRSST-PP GDSv1 metadata records

This appendix to the GDSv1 describes the GHRSST-PP Master Metadata Repository (MMR) service, the format of GHRSST-PP metadata records that are generated for each GDSv1 data file and how these are delivered, registered, modified and retrieved from the MMR. The MMR also serves the GHRSST-PP high resolution diagnostic data set (HR-DDS) system.

A6.1 The GHRSST-PP Master Metadata Repository (MMR) system

In order to successfully exchange and manipulate data in a real time environment, RDAC and GDAC must know at any instant what data are available, where they are and what level of processing they have completed. This is a particularly demanding task given the global distribution of RDAC and GDAC that are working together. In order to address these issues, a global GHRSST-PP Master Metadata Repository service has been developed. The MMR is required to ensure that all data resources are "visible" to the RDAC and GDAC centres and it is the "core" of the GHRSST-PP data management framework. It provides a searchable catalogue of the distributed GHRSST-PP data holdings providing information on its physical location, contents and any constraints on its use. Without a master catalogue, it would be extremely difficult (if not impossible) to locate a single data resource within the GHRSST-PP without physically connecting to the computers storing the data and searching each one individually.

The MMR is a physical database containing metadata descriptions of all transient and non-transient data holdings within the GHRSST-PP. A minimum set of geo-spatial information describing a particular data granule (in situ observation, satellite image) that has been used in the GDSv1 will be automatically stored within MMR database system in real time. The GHRSST-PP MMR data record format is based on the Global Change Master Directory (GCMD) Directory Interchange Format (DIF) standard. The GCMD has an extensive number of keyword lists that can be used for encoding the physical quantities, the data centres, *etc.* Use of these keyword lists provides a simple and efficient way for users to make unambiguous searches of the database in order to locate the data they want. GCMD also has a large set of sensibly defined metadata fields based on many years of experience of serving metadata.

The GHRSST-PP data files themselves have been chosen to follow the Climate and Forecast netCDF conventions (See Appendix 1) because these conventions provide a practical standard for storing oceanographic data, and have already been adopted for the Data Sharing Pilot project within GODAE. The global attributes of netCDF files are also metadata, since they follow the CF conventions, although they differ slightly from the GHRSST-PP MMR standard. To a certain extent, this is reasonable, since the MMR is intended to help users and processing centres search for data sets and to keep track of processing tasks, while the netCDF attributes are to help users apply the data.

The following section describes the format of GHRSST-PP metadata records. The GHRSST-PP MMR system is fully documented in the GDIP.

A6.2 GHRSST-PP Master Metadata Repository metadata record format

The metadata standard defined for GHRSST-PP MMR system is based on the GCMD DIF standard. The following web reference documents are applicable to GDSv1 metadata records:

GCMD DIF definition document http://gcmd.gsfc.nasa.gov

GHRSST-PP metadata consist of two related but distinct types of metadata record:

1. A Data Set Record (MMR-DSR) contains information which is common to all files within a data set, such as the title and contact information and a summary of the data set contents. There is one, and only one, MMR-DSR metadata record for each data set. A DSR exists for each GDSv1 data product type at each RDAC and GDAC. In addition, a DSR exists for every GHRSST-PP High Resolution Diagnostic Data Set (HR-DDS) site as explained in Appendix 5. The format of MMR-DSR records are given in Table A6.2.1.
2. Each file in a data set described by a MMR-DSR is represented by a unique File Record (MMR-FR) metadata record. A MMR-FR metadata record contains information that distinguishes that file from all others in the data set. The format of MMR-FR is given in Table A6.2.2. There are many different MMR-FR metadata records associated with a single MMR-DSR metadata record.

Within the GDSv1 only MMR-FR metadata records need to be created and submitted to the MMR system operationally by RDAC and GDAC as DSR metadata records will already be present within the MMR system. Every time a GDS data file or is produced, an associated MMR-FR metadata record must be created (see Appendix A6.3) and registered at the MMR via e-mail (see Section A6.4).

Table A6.2.1. Structure template for a GHRSST-PP metadata Data Set Record (MMR-DSD, v3.0a, 21/10/2002). Mandatory fields are shown in blue highlight.

Field Name	Type	Description
`Entry_ID`	Char[160]	A unique identifier for the data set. . The indentifier will be set by the project and reflect the location, source, sensor, parameter and product of the data set, e.g., USA-RSS-AMSRE-MW-L2-SST (AMSR-E microwave L2 SST data from RSS , Inc. in the USA)
`Entry_Title`	Char[160]	Data set title • Title should be descriptive enough so that when a user is presented with a long list of titles, the user can determine the general content of the data set. • In order to make titles descriptive, important elements about the data may be included in the title, *i.e.,* parameters measured, geographic location, instrument, investigator, project, temporal coverage. • For readability, capitalization of the title should follow standard constructs. Do not use all capital letters or all lower case letters, but use the appropriate case where applicable. Example "Sea Surface Temperature from the Along-Track Scanning Radiometer 2 (ATSR-2), 1km resolution, daily, global, for the year 1999, from the Rutherford Appleton Laboratory, UK"
`Data_Set_URL`	Char[512]	The URL providing direct manual FTP access to the directory containing the dataset, for those wishing to download the entire data set.
`Group: Dataset Citation`		A citation for the data set to properly credit the data set producer. This field has two functions: 1. to indicate how this data set should be cited in the professional scientific literature, and 2. if this data set is a compilation of other data sets, to document and credit the data sets that were used in producing this compilation. This group is not to be used to list bibliographic references of scientific research articles arising from the data set. This field provides a citation for the data set itself, not articles related to the research results.
`Dataset_Creator`	Char[160]	The name of the organization(s) or individual(s) with primary intellectual responsibility for the data set's development
`Dataset_Title`	Char[160]	The title of the data set; this may be the same as Entry Title
`Dataset_Release_Date`	Char[80]	The date when the data set was made available for release
`Dataset_Release_Place`	Char[80]	The name of the city (and state or province and country if needed) where the data set was made available for release
`Dataset_Publisher`	Char[80]	The name of the individual or organization that made the data set available for release

Version	Char[80]	The version of the data set
Issue_Identification	Char[80]	The volume or issue number of the publication
Data_Presentation_Form	Char[80]	The mode in which the data are represented, *e.g.* atlas, image, profile, etc
Other_Citation_Details	Char[160]	Any other details
Online_Resource	Char[160]	The URL of an online computer resource containing user-oriented information about the data set (e.g. help files, articles), if it exists.
Group: Parameters		This group describes the types of measurements represented by the data, such as sea surface temperature, windspeed, *etc*.
Category_Topic_Term	Keyword	This is a concatenation of the three GCMD DIF keyword fields: Category, Topic and Term, which cover a fairly exhaustive range of data types. Units should be stated in the data files themselves
Variable	Keyword	Another keyword provding further specification of the measurement type
Detailed_Variable	Char[80]	A free text field for provding further specification of the measurement type
Sensor_Name	Keyword	The instrument or hardware used to acquire the data
Access_Constraints	Char[1024]	Restrictions, limitations and legal prerequisites for accessing the data set. Some words which may be used in this field include: Public, In-house, Limited, Source. e.g. "Data available to the general public after 5 year embargo period beginning January 1, 1994"
Use_Constraints	Char[1024]	Restrictions, limitations and legal prerequisites for using the data set. e.g. "Data may not be used for commercial applications".
Group: Data Center		This group identifies the data centre which distributes the data, the data centre URL, data set identification, and a person to contact. Note USA spelling of centre
Data_Center_Name	Keyword	Name of the data centre taken from the list of valid keywords. Data centres which are part of a larger organization may have a compound short name made up of the organization short name and the data center short name; for example: INPE/DAE/FISAT. This is helpful for users conducting broad searches because the users may query on the substring INPE and receive all metadata records from INPE, regardless of division.
Data_Center_URL	Char[160]	The Internet Uniform Resource Locator(s) (URL) for the data centre should be listed if available. This field may be repeated as many times as necessary within the data centre group. The URLs will be hypertext linked. The URLs may include gopher, ftp and telnet as well as WWW servers e.g., • Data_Center_URL: http://daac.gsfc.nasa.gov • Data_Center_URL: ftp://podaac.jpl.nasa.gov • Data_Center_URL: telnet://ncar.ucar.edu • Data_Center_URL: gopher://www.ciesin.org
Data_Set_ID	Char[160]	Should be listed if available. These are identification codes assigned by the data centre, which will simplify the location and ordering of the data sets.
Group: Personnel		This group describes the primary contact person at the data centre who is able to respond

		to requests for and queries about the data. Most fields are self-explanatory. Address should be a valid postal address for the named person.
First_Name	Char[80]	
Middle_Name	Char[80]	
Last_Name	Char[80]	
Email	Char[80]	
Phone	Char[80]	
Fax	Char[80]	
Address	Char[1024]	

Summary	Char[2048]	A brief description of the data set, descriptive enough to allow potential users of the data set to determine if the data set is useful for their needs.
		• Should include information needed for a user to determine the usefulness of the data set.
		• Should start with a topic sentence, describing what information is in the data set. Often, this is some measurable quantity or quantities, such as sea surface temperature, human population density, or species mortality rate.
		• The total length should consist of approximately 30 lines, with each line not exceeding 80 characters, and separated by a carriage return (or carriage return + line feed) character.
		• Should restate information that may be found in other fields if that information is vital to the understanding of the data set (*e.g.*, parameters, spatial coverage).
		• Should reference the source information if the summary was abstracted from an existing document.
		• Single spaced with blank lines separating paragraphs
		• Capitalization should follow standard constructs. For readability, do not use all capital letters or all lower case letters, but use the appropriate case where applicable.
		• No right justification
		• Acronyms should be expanded to improve understanding.
		• May contain tabular information
		• Hyperlinked URLs may be embedded in the text by surrounding them with single or double quotes: "http://www.ngdc.noaa.gov"
		• Where applicable, should include brief statements of the following important information:
		    ○ Data processing information (gridded, binned, swath, raw, algorithms used, necessary ancillary data sets)
		    ○ Methodology or analytical tools

		o Time gaps in data set coverage o Units and unit resolution o Similarities and differences of these data to other closely-related data sets o Other pertinent information
Location	Keyword	Location of the data set (e,g., Atlantic Ocean)
Metadata_Version	Char[31]	A number providing the metadata version, e.g. "2.1". Only valid version numbers should be used - check with the www.ghrsst-pp.org web site.
Metadata_Creation_Date	Char[10]	The date the metadata was created. Years should be four-digits. Month and day of month should be two digits, with leading zeroes if necessary
Metadata_Last_Revision_Date	Char[10]	The date the metadata was last revised. Format must be "yyyy-mm-dd". Years should be four-digits. Month and day of month should be two digits, with leading zeroes if necessary

Table A6.2.2. Structure template for a GHRSST-PP metadata File Record (MMR-FR, v3.0a, 21/10/2002). Mandatory fields are shown in light blue highlights

Field Name	**Type**	Description
Entry_ID	Char[160]	Filename suffix appended to Entry_ID for commensurate data set metadata entry, e.g. USA-RSS-AMSRE-MW-L2-SST.2004181.2315.amsr-e.sst.ncdf. There may be additional number after this reflect differing algorithms or other factors that could be assigned on case-by-case basis.
Data_File_URL	Char[512]	This is a URL providing direct access to the GHRSST-PP data file
Group: Temporal Coverage		*This group* provides the start and end date and times for the measurements contained in the GHRSST-PP data file. All date and times must be in Universal Time Coordinates (UTC, which is approx. Greenwich Mean Time), and formatted at "yyyy-mm-dd" for dates and "hh:mm:ss" for times.
Start_Date	Char[10]	
Start_Time	Char[8]	
Stop_Date	Char[10]	
Stop_Time	Char[8]	
Group: Spatial Coverage		Geographic coverage (horizontal and vertical) of the data described. Longitude must be expressed in the range -180.0 to +360.0 as an offset from the Greenwich meridian, and latitude in the range -90.0 to +90.0, as an offset from the Equator. For gridded data, these coordinates should be the outer edges of the coverage, and not the coordinates of the centres of the edge pixels.
Southernmost_Latitude	Float	
Northernmost_Latitude	Float	

Westernmost_Longitude	Float	
Easternmost_Longitude	Float	
Minumum_Altitude	Char[80]	the altitude level which represents the lower limit of data coverage, as measured from mean sea level
Maximum_Altitude	Char[80]	the altitude level which represents the higher limit of data coverage, as measured from mean sea level.
Minumum_Depth	Char[80]	the depth level which represents the upper-most depth of data coverage, as measured from mean sea level.
Maximum_Depth	Char[80]	the depth level which represents the lowest depth of data coverage, as measured from mean sea level.
Group: Data Resolution		
Spatial_Resolution	Char[80]	
Temporal_Resolution	Char[80]	
Altitude_Resolution	Char[80]	
Depth_Resolution	Char[80]	
Group: Projection_Information		
Projection_Type	Keyword	A keyword from the GHRSST-PP approved list of projection names.
Ellipsoid_Type	Keyword	A keyword from the GHRSST-PP approved list of ellipsoids.
Other_Projection_Details	Char[160]	Free text description of details such as the projections standard parallels, *etc.*, if necessary to fully specify the projection.
Metadata_Creation_Date	Char[10]	The date the metadata was created. Years should be four-digits. Month and day of month should be two digits, with leading zeroes if necessary
Metadata_Last_Revision_Date	Char[10]	The date the metadata was last revised. Format must be "yyyy-mm-dd". Years should be four-digits. Month and day of month should be two digits, with leading zeroes if necessary

A6.3 XML Document Type Definition (DTD) for MMR-DSR and MMR-FR metadata records.

All GHRSST-PP MMR metadata records are encoded using the extensible mark-up language (XML). XML provides an application independent way of sharing and controlling the quality and content of data. A detailed description of XML-v1.0 can be found at http://www.w3.org/TR/REC-xml

An XML Document Type Definition (DTD) is a formal description in XML Declaration Syntax of a particular type of document. A DTD provides applications with advance notice of what names and structures can be used in a particular document type. Using a DTD when editing files means you can be certain that all documents which belong to a particular type will be constructed and named in a consistent and conformant manner. Validating XML parsers read a DTD before they read an XML document so that they can identify where every element type ought to come and how each element relates to the other. A major advantage if using an XML DTD is that data records can be validated providing a quality control method well suited to database applications such as the MMR that are expecting a data record in a specific format.

A DTD can be declared inline within an XML document itself, or as an external reference. The MMR uses a GHRSST-PP XML DTD to verify that MMR metadata records received from the GHRSST-PP RDAC and GDAC are valid. MMR DTD can also be used to validate MMR data records at the time of metadata record creation.

Figure A6.3.1 provides an XML DTD for MMR-DSD v3.0a metadata records that are described in Table A6. 2.1

Figure A6.3.1 XML DSD for MMR-DSR records (MMR-DSD, v3.0a, 21/10/2002).

```
<?xml version="1.0"?>
<!—DTD for MMR-DSD v3.0a 21 October 2002-->
<!—C J Donlon -->
<!DOCTYPE MMR-DSD [
     <!ELEMENT DSD_Entry_ID (#PCDATA)>
     <!ELEMENT Data_Creator (#CDATA)>
     <!ELEMENT Data_Title (#CDATA)>
     <!ELEMENT Data_Release_Date (#PCDATA)>
     <!ELEMENT Version (#PCDATA)>
     <!ELEMENT Issue_Identification(#PCDATA)>
     <!ELEMENT Data_Presentation_Form (#PCDATA)>
     <!ELEMENT Other_Citation_Details (#CDATA)>
     <!ELEMENT Variable(#CDATA)>
     <!ELEMENT Access_Constraints (#CDATA)>
     <!ELEMENT Use_Constraints (#CDATA)>
     <!ELEMENT Data_Center_Name (#PCDATA)>
      <!ELEMENT Data_Center_URL (#PCDATA)>
      <!ELEMENT Data_Set_ID (#PCDATA)>
      <!ELEMENT First_Name (#CDATA)>
      <!ELEMENT Middle_Name (#CDATA)>
      <!ELEMENT Last_Name (#CDATA)>
      <!ELEMENT Email (#CDATA)>
      <!ELEMENT Phone (#CDATA)>
      <!ELEMENT Fax (#CDATA)>
     <!ELEMENT Address (#CDATA)>
     <!ELEMENT Summary (#CDATA)>
     <!ELEMENT Location (#CDATA)>
     <!ELEMENT Metadata_version (#PCDATA)>
```

```
        <!ELEMENT Metadata_Creation_Date (#PCDATA)>
        <!ELEMENT Metadata_Last_Revision_Date (#PCDATA)>
    ]
```

Notes:

- <!ELEMENT element-name (#PCDATA)> means that the element contains data that is going to be parsed by the MMR parser.
- <!ELEMENT element-name (#CDATA)> means that the element contains data that is not going to be parsed by the MMR parser.

Figure A6.3.2 provides an XML DTD for MMR-FR v3.0a metadata records that are described in Table A6. 2.1. The additional element `Update_Metadata_entry` is used to tell the MMR system that this metadata record should replace any existing data in the database.

Figure A6.3.2 XML DSD for MMR-FR records (MMR-FR, v3.0a, 21/10/2002).

```
<?xml version="1.0"?>
<!—DTD for MMR-FR v3.0a 21 October 2002-->
<!—C J Donlon -->
<!DOCTYPE MMR-FR [
     <!ELEMENT DSR_Entry_ID (#PCDATA)>
     <!ELEMENT Data_File_URL (#PCDATA)>
     <!ELEMENT Start_Date (#PCDATA)>
     <!ELEMENT Start_Time (#PCDATA)>
     <!ELEMENT Stop_Date (#PCDATA)>
     <!ELEMENT Stop_Time (#PCDATA)>
     <!ELEMENT Southernmost_Latitude (#PCDATA)>
     <!ELEMENT Northernmost_Latitude (#PCDATA)>
     <!ELEMENT Westernmost_Longitude (#PCDATA)>
     <!ELEMENT Easternmost_Longitude (#PCDATA)>
     <!ELEMENT Minumum_Altitude (#PCDATA)>
     <!ELEMENT Maximum_Altitude (#PCDATA)>
     <!ELEMENT Minumum_Depth (#PCDATA)>
     <!ELEMENT Maximum_Depth (#PCDATA)>
     <!ELEMENT Spatial_Resolution (#PCDATA)>
     <!ELEMENT Temporal_Resolution (#PCDATA)>
     <!ELEMENT Altitude_Resolution (#PCDATA)>
     <!ELEMENT Depth_Resolution (#PCDATA)>
     <!ELEMENT Projection_Type (#PCDATA)>
     <!ELEMENT Ellipsoid_Type (#CDATA)>
     <!ELEMENT Other_Projection_Details (#PCDATA)>
     <!ELEMENT Metadata_record_issue (#PCDATA)>
     <!ELEMENT Update_Metadata_entry (#PCDATA)>
     <!ELEMENT Metadata_version (#PCDATA)>
     <!ELEMENT Metadata_Creation_Date (#PCDATA)>
     <!ELEMENT Metadata_Last_Revision_Date (#PCDATA)>
]>
```

A6.4 Delivery and registration of MMR-FR metadata to the MMR

In order for the MMR system to ingest and register GHRSST-PP MMR-FR metadata records, it is necessary to

1. Generate appropriate MMR-FR element data according to Table A6. 2.2
2. Format the MMR-FR as an XML document according to the XML DTD specified in Figure A6.3.2

3. Validate the XML MMR-FR XML document using the XML DTD specified in Figure A6.3.2.
4. Send the MMR-FR XML document to the MMR

Each of these different procedures is explained in the following paragraphs.

A6.4.1 Preparing an GDSv1 MMR-FR metadata record

Each metadata record should be encoded in extensible mark-up language (XML) ASCII text format. Figure A6.4.1 provides an example MMR-FR metadata record formatted as an XML document according to the DTD provided in Figure A6.2.2. The metadata record describes a L2 AMSRE data file produced at the USA RDAC.

Figure 6.4.1 Example XML metadata record defined in Table A6.2.2 and formatted according to the XML DTD described in Figure A6.3.2

```
<?xml version='1.0' encoding='utf-8'?>
<!--  MR-FR example for MMF-FR v3.0a v0.1 -->
<MMR-FR
DSR_Entry_ID=" USA-RSS-AMSRE-MW-L2-SST"
     Data_File_URL="ftp://www.mydatacentre.com/ghrsst_data/RSS-
     AMSRE-MW-L2-SST.2004181.2315.amsr-e.sst.ncdf"
     Start_Date="03-06-21"
     Start_Time="12:00:00"
     Stop_Date="03-06-21"
     Stop_Time="18:00:00"
     Southernmost_Latitude="-27.223"
     Northernmost_Latitude="56.887"
     Westernmost_Longitude="23.3345"
     Easternmost_Longitude="-1.334"
     Minumum_Altitude="0"
     Maximum_Altitude="0"
     Minumum_Depth="0"
     Maximum_Depth="0"
     Spatial_Resolution="9.28"
     Temporal_Resolution="6"
     Altitude_Resolution="0"
     Depth_Resolution="0"
     Projection_Type="Pathfinder Equal Area"
     Ellipsoid_Type=""
     Other_Projection_Details=""
     Metadata_record_issue="1"
     Update_Metadata_entry="No"
     Metadata_version="ghrsst-pp-v0.1"
     Metadata_Creation_Date="03-01-23"
     Metadata_Last_Revision_Date="03-01-21"
/>
```

A6.4.2 Validating an GDSv1 MMR-FR metadata record

There are a number of automated quality control checks that are being implemented,: all the required fields need to be there with valid entries.

A6.4.3 Registering an GDSv1 MMR-FR metadata record at the MMR

GDSv1 MMR-FR are registered with the GHRSST-PP MMR system via e-mail. Each XML encoded MMR-FR is entered as the body of an ASCII text e-mail message which is then sent to:

MMR_ghrsst@podaac.jpl.gov

The subject line of each metadata message should be formatted as follows:

```
Subject: GHRSST metadata notification from <processing
centre code>
```

where <processing centre code> is defined in Table 2.7.1. For example

```
Subject: GHRSST metadata notification from JAP
```

would indicate that the e-mail contains a metadata record from the Japanese RDAC system. Figure 6.4.3.1 shows an example metadata delivery message header.

Figure A6.4.2.1 An example e-mail header fro a MMR-FR delivery message.

```
Subject: GHRSST metadata notification from EUR
Date: Thu, 30 Jan 2003 17:42:17 -0800
From: Jorge Vazquez <jv@pacific.jpl.nasa.gov>
Organization: JPL/Caltech
To: GHRSST-PP MMR <MMR_ghrsst@podaac.jpl.gov>
References: <3E385EC5.18CF9B3F@seanet.jpl.nasa.gov>
```

If the MMR XML parser rejects the MMR-FR metadata message, it will return an e-mail with a full diagnostic report explaining the cause of the error to the return address specified in the delivery e-mail. A revised metadata message should be prepared and sent to the MMR as soon as possible

A6.4.4 Modifying GDSv1 metadata records already registered at the MMR

If an incorrect MMR-FR metadata record is sent to the MMR or an MMR-FR needs to be revised, a newm etadata message is sent to the MMR. The MMR-FR element `metadata_record_issue` should be incremented and the MMR-FR element `Update_Metadata_Entry` field set to “yes” as shown in Figure A6.4.4.1.

Note that the `Update_Metadata_entry=”yes”` entry indicates that this metadata record should replace any existing data already registered at the MMR under this `DSR_Entry_ID`. **CAUTION: This action will replace any previous version of the metadata record held in the MMR database system with the revised version.**

Figure 6.4.4.1 Example XML metadata record defined in Table A6.2.2 and formatted according to the XML DTD described in Figure A6.3.2

```
<?xml version='1.0' encoding='utf-8'?>
<!--  MR-FR example for MMF-FR v3.0a v0.1 -->
<MMR-FR
     DSR_Entry_ID=” USA-RSS-AMSRE-MW-L2-SST”
```

Data_File_URL="ftp://www.mydatacentre.com/ghrsst_data/RSS-AMSRE-MW-L2-SST.2004181.2315.amsr-e.sst.ncdf"
Start_Date="03-06-21"
Start_Time="12:00:00"
Stop_Date="03-06-21"
Stop_Time="18:00:00"
Southernmost_Latitude="-27.223"
Northernmost_Latitude="56.887"
Westernmost_Longitude="23.3345"
Easternmost_Longitude="-1.334"
Minumum_Altitude="0"
Maximum_Altitude="0"
Minumum_Depth="0"
Maximum_Depth="0"
Spatial_Resolution="9.28"
Temporal_Resolution="6"
Altitude_Resolution="0"
Depth_Resolution="0"
Projection_Type="Pathfinder Equal Area"
Ellipsoid_Type=""
Other_Projection_Details=""
Metadata_record_issue="1"
Update_Metadata_entry="yes"
Metadata_version="ghrsst-pp-v0.1"
Metadata_Creation_Date="03-01-23"
Metadata_Last_Revision_Date="03-01-21"
/>

A6.5 Retrieval of metadata from the MMR

MMR data may be retrieved vi a web browser interface.

Appendix 7. GDSv1 processing system logs

A7.1 The ISDI Operational Log (OPLOG)

The GDSv1 operational log is a central e-mail managed system that collates and publishes all of the operations within the GDSv1. Each time an operation is started or competed at an RDAC or GDAC an e-mail is sent to the GDS operations log OPLOG@ghrsst-pp.org and the mail is automatically published at an OPLOG web page located at http://www.ghrsst-pp.org/ISDIv1OPS. Table A7.1 describes the events that should generate an OPLOG message.

Table 7.1 Significant GDSv1 events that generate an OPLOG message.

Event code	Description
1	RDAC centre is up
2	RDAC/GDAC centre is down
3	Collated data product sent to GDAC
4	Merged data product sent to GDAC
5	L2P data processing synopsis report (L2P_DPSR)
6	L4 data processing synopsis report L4_DPSR
7	…

DPSR=data processing synopsis report

A7.2 The ISDI Error Log (ERRLOG)

The GDSv1 error log is a central e-mail managed system that collates and publishes all of the errors that have occurred within the GDS distributed processing system. Each time an error is raised at an RDAC or GDAC an e-mail is sent to the GDS error log ERRLOG@ghrsst-pp.org and the mail is automatically published at an ERRLOG web page located at http://www.ghrsst-pp.org/ISDIv1ERR.

Appendix 8 The GDSv1 Problem resolution board

"The Problem Resolution Board" provides a traceable mechanism for decisions that are required to improve problematic situations or errors. The request for a solution may be passed on to a recognized expert who will provide a recommended course of action in a timely manner. The use of a Board to resolve problems helps avoid a solution being implemented but in so doing unintentionally creating another problem through being unaware of downstream implications.

The Problem Resolution Board (PRB) is based on the ARM approach. There may be other, possibly better, ways of solving GDSv1 problems but this approach does provide traceability of decisions, and helps avoid someone making a change to solve one problem but in so doing unintentionally creating another one through being unaware of downstream implications.

The GDS problem resolution board should contain the following sub-groups:

- **Cal-Val problem resolution board**
- **Data management problem resolution board**
- **GDS problem resolution board**
- **ISDI-OPS and ERRLOG log monitors**

The Board and sub-groups should meet on a regular and frequent basis, by email or teleconference

As a general guide, for each problem situation, the PRB must establish:

- What is the control loop to decide on what to do?
- How should the ISDI-TAG deal with this?
- What are the allowable timescales?
- How should operational users be informed of the problem and how quickly can they be informed and by which route (e-mail??)
- What is the process for rectifying anomalies?
- What is the control loop to decide on what to do?
- How should the ISDI-TAG deal with this?
- What are the allowable timescales?
- How should operational users be informed of the problem and how quickly can they be informed and by which route (e-mail??)

The GHRSST-PP GDS PRB will be discussed at the GHRSST-PP workshop, September 2003.